University of Cape Town

Faculty of Health Sciences

Division of Biomedical Engineering

Reconstruction of three-dimensional facial geometric features related to fetal alcohol syndrome using adult surrogates

Felix Atuhaire (ATHFEL001)

Supervisor

Dr. Tinashe Mutsvangwa

Co-supervisor

Professor Tania Douglas

A dissertation submitted in fulfillment of the requirements for the degree of MSc in Biomedical Engineering by research

August 31, 2020

Declaration

I, **Felix Atuhaire**, hereby declare that the work in this dissertation for a Master of Science Degree in Biomedical Engineering is a compilation of my original work, except where references or acknowledgements indicate otherwise. I also confirm that the work this dissertation has been compiled entirely by myself, been solely the result of my research study, and not been submitted for any other degree or professional qualification in this or another university. I allow the university to reproduce my work for purposes of research either as the whole or any portion of the contents in whatsoever manner.

Signed: Signed by candidate

Date: August 31, 2020

Abstract

Fetal alcohol syndrome (FAS) is a condition caused by prenatal alcohol exposure. The diagnosis of FAS is based on the presence of central nervous system impairments, evidence of growth abnormalities and abnormal facial features. Direct anthropometry has traditionally been used to obtain facial data to assess the FAS facial features. Research efforts have focused on indirect anthropometry such as 3D surface imaging systems to collect facial data for facial analysis. However, 3D surface imaging systems are costly. As an alternative, approaches for 3D reconstruction from a single 2D image of the face using a 3D morphable model (3DMM) were explored in this research study.

The research project was accomplished in several steps. 3D facial data were obtained from the publicly available BU-3DFE database, developed by the State University of New York. The 3D face scans in the training set were landmarked by different observers. The reliability and precision in selecting 3D landmarks were evaluated. The intraclass correlation coefficients for intra- and inter-observer reliability were greater than 0.95. The average intra-observer error was 0.26 mm and the average inter-observer error was 0.89 mm. A rigid registration was performed on the 3D face scans in the training set. Following rigid registration, a dense point-to-point correspondence across a set of aligned face scans was computed using the Gaussian process model fitting approach. A 3DMM of the face was constructed from the fully registered 3D face scans. The constructed 3DMM of the face was evaluated based on generalization, specificity, and compactness. The quantitative evaluations show that the constructed 3DMM achieves reliable results. 3D face reconstructions from single 2D images were estimated based on the 3DMM. The Metropolis-Hastings algorithm was used to fit the 3DMM features to 2D image features to generate the 3D face reconstruction. Finally, the geometric accuracy of the reconstructed 3D faces was evaluated based on ground-truth 3D face scans. The average root mean square error for the surface-to-surface comparisons between the reconstructed faces and the ground-truth face scans was 2.99 mm.

In conclusion, a framework to estimate 3D face reconstructions from single 2D facial images was developed and the reconstruction errors were evaluated. The geometric accuracy of the 3D face reconstructions was comparable to that found in the literature. However, future work should consider minimizing reconstruction errors to acceptable clinical standards in order for the framework to be useful for 3D-from-2D reconstruction in general, and also for developing FAS applications. Finally, future work should consider estimating a 3D face using multi-view 2D images to increase the information available for 3D-from-2D reconstruction.

Acknowledgements

I would like to express my sincere appreciation to my supervisor, Dr Tinashe Mutsvangwa, for the continuous support during the research project and the writing of my master's dissertation; for his patience, guidance, motivation to get the work done; and the immense knowledge that he shared with me. I could not have imagined a better supervisor and mentor for my master's degree.

I would also like to thank Professor Tania Douglas, who was my co-supervisor for the support, guidance, reviewing of my work, and facilitating the financial support for my master's degree.

My sincere thanks go out to my fellow group mates in the Medical Image Inferencing and Distributed Diagnostics (MI2D2) Group including: Jean-Rassaire Fouefack, Sarah McClean, Khwezi Majola, Catherine Namayega, Yvonne Karanja, Xolisile Thusini, Tewodros Belay, and Herve Nicholas for the fun, support, encouragement, life stories, and resilience during the period of the research project.

Special thanks to the African Biomedical Engineering Mobility (ABEM) project, under the Intra-Africa Academic Mobility Scheme of the European Commission's Education, Audiovisual and Culture Executive Agency as well as the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation in South Africa (Grant no. 98788) for funding my master's research.

The research project computations were performed using facilities provided by the University of Cape Town's Information and Communication Technology Services (ICTS) High Performance Computing team.

Finally, I wish to thank my parents George W M and Geraldine K for the gift of giving birth to me, supporting and protecting me throughout my growth to now being "a big" man. May you live longer.

List of abbreviations

SSMs:	Statistical shape models
ICP:	Iterative closest point
3DMMs:	Three-dimensional morphable models
FAS:	Fetal alcohol syndrome
2D:	Two-dimensional
3D:	Three-dimensional
GPA:	General Procrustes analysis
GPs:	Gaussian processes
GPMMs:	Gaussian process morphable models
SAMs:	Statistical appearance models
PCA:	Principal component analysis
BFM:	Basel face model
GDA:	Generalized discriminant analysis
NICP:	Non-rigid iterative closest point
LSFM:	Large scale facial model
ANOVA:	Analysis of variance
CI:	Confidence interval
CT:	Computed tomography
ICC:	Intraclass correlation coefficient

Table of Contents

List of figures

List of Tables

Chapter 1. Introduction

1.1 Background

Fetal alcohol syndrome (FAS) is a non-hereditary condition which presents as birth defects and developmental disabilities in affected individuals (Jones, 2014). This condition is a result of prenatal alcohol exposure. The global prevalence of FAS is estimated at 14.6 per 10,000 population (Popova *et al.*, 2017). In Sub-Saharan Africa, South Africa has the highest prevalence rate of FAS between 55.42 to 79 per 1,000 live births (May *et al.*, 2016; Roozen *et al.*, 2016). The consistent diagnostic features of the FAS include growth deficiencies (decreased weight and height), impairments of the central nervous system (small head circumference or brain malformations), and distinctive facial features (Astley *et al.*, 1995, 2000; Moore *et al.*, 2007). The distinctive facial features of FAS can be classified as either discriminative or associative (Sampson *et al.*, 1997). The associative facial features of FAS include low nasal bridge, epicanthal folds, micrognathia, minor ear anomalies, and short noses. The discriminative facial features of FAS, which uniquely characterize the syndrome, include a thin upper lip, short palpebral fissure lengths, and a smooth philtrum (Astley *et al.*, 1995, 2000; Astley *et al.*, 1999; Sampson *et al.*, 1997). Studies have indicated that physical facial features of FAS tend to disappear or become less prominent in individuals as they develop into adults especially over 12 years (Huang *et al.*, 2005; Mutsvangwa *et al.*, 2010; Streissguth *et al.*, 1992).

Aside from the common characteristics of FAS, individuals with FAS tend to develop several neurobehavioral challenges including depression, anxiety, and acute stress disorder at later stages of their life (Streissguth *et al.*, 2004). It is believed that early detection of FAS is vital for providing early interventions that could reduce the onset of secondary disorders. Secondary disorders include mental breakdown, improper sexual behaviors, and disrupted school experience (Streissguth *et al.*, 1996; Streissguth *et al.*, 2004). Furthermore, early diagnosis of FAS could be beneficial to the birth mothers by providing interventions (such as education and sensitization) that may prevent the birth of children with FAS in the future (Astley *et al.*, 2000).

To assess the physical characteristics of FAS, facial measurements are carefully acquired using data collection methods which have been proposed in the literature. These data collection methods include direct anthropometry such as calipers and indirect anthropometry such as two-dimensional (2D) photogrammetry, three dimensional (3D) stereo-photogrammetry, and 3D surface imaging (Astley *et al.*, 1995; Chang *et al.*, 2015; Honrado *et al.*, 2004; Lekakis *et al.*, 2016; Meintjes *et al.*, 2002; Moore *et al.*, 2001). Compared to the indirect anthropometry, direct anthropometry introduces inaccuracies due to indentation of some features during contact measurements using physical instruments. The indirect anthropometry methods obtain photographs or 3D data within less time compared to taking measurements directly and measurements on the images can be repeated in the absence of subjects. While 3D surface imaging system for indirect anthropometry has been proposed by Hammond *et al.* (2004), surface imaging systems tend to be costly and therefore inaccessible especially in low resource settings. Existing statistical approaches that estimate 3D surfaces of objects from single 2D images could serve as alternative solutions for obtaining facial data for the analysis of the FAS facial phenotype.

Information extracted from facial images may be used to explain similarities and differences existing within or between populations. For example, the features extracted from facial images may be used for explaining medical conditions (Schonborn *et al.*, 2017) and performing morphometric studies (Bacivarov *et al.*, 2009; Mutsvangwa *et al.*, 2007; Mutsvangwa *et al.*, 2010). Face analysis approaches may be aided using tools such as three dimensional morphable models (3DMMs).

Three-dimensional morphable models (3DMMs) are high resolution statistical models containing shape and texture variability information from a sample population (Blanz *et al.*, 1999; Booth *et al.*, 2017; Booth *et al.*, 2018; Dai *et al.*, 2017). The 3DMMs may be used as parametric generative models as explained in the seminal work of Blanz *et al.* (1999). Typically, 3DMMs are constructed from a collection of 3D face scans after establishing anatomical correspondences across the scans in the dataset. Establishing correspondences ensures that similar features across the collection of 3D face scans match each other. The 3DMMs may be categorized into statistical shape models (SSMs) or statistical shape and appearance models (SSAMs). As the name suggests, SSMs encode shape variability only, while SSAMs encode both shape and texture variability. The applications

of 3DMMs of the face include face reconstruction, face recognition, face animation, face tracking, and face surface segmentation. The coupling of the 3DMMs of the face with mobile phone-based platforms has the potential to provide a useful computer-assisted adjunct for screening and diagnosis of facial conditions related to FAS.

The fitting methods for approximating a 3D face reconstruction from a single 2D image of the face using a 3DMM have been reported in the literature. For instance, research by Blanz *et al.* (1999) demonstrated a fitting approach for approximating a 3D face (shape and texture) from a single 2D image of the face based on a 3DMM. These fitting methods use 3DMMs as shape and texture priors to estimate a 3D human face from one or multiple 2D images. The fitting methods provide alternative approaches compared to using direct 3D surface imaging systems to obtain 3D facial geometries for facial analysis. The fitting methods require 2D facial images as inputs and 3DMMs working as prior knowledge. The 2D facial images can be conveniently captured with either a smartphone camera or a digital camera. An additional benefit is the short acquisition time required to capture 2D facial images compared to capturing 3D face scans. However, current applications of these fitting methods have not quantified the geometric accuracy of the resultant 3D face. Moreover, it is not known whether these methods can capture the subtle geometric deformations that occur in human faces of individuals with FAS. This project explored the potential of reconstructing a 3D human face from a single 2D image based on a 3DMM. Furthermore, an assessment of the geometric accuracy of the reconstructed 3D faces, targeting facial features associated with FAS, was performed and is reported.

1.2 Aim and objectives

This project aimed to develop and validate a framework for estimating a 3D reconstruction of the human face from a single 2D image of the face using a 3D morphable model as a prior. The following objectives were identified as the path to achieving the aim:

1. Develop and evaluate the performance of a 3D morphable model of the human face.
2. Estimate a 3D reconstruction of the human face from a single 2D image of a face using a 3D morphable model.

3. Evaluate the geometric accuracy of 3D reconstruction of the human face considering the facial features associated with FAS diagnosis.

1.3 Limitations of the project study

The research project used image data from a healthy adult population for developing and evaluating the reconstruction framework based on the 3D morphable model of the face. The rationale was that if the framework provides accurate 3D face reconstructions using the normal facial data, then the same framework could be used in future to estimate the 3D faces of individuals with deformities related to FAS provided the 3D morphable model of such individuals was used as the shape prior. Additionally, this research project did not attempt to classify human faces of individuals with FAS, which would have required datasets of faces of individuals with and without FAS.

1.4 Overview of the dissertation

The dissertation is structured as follows. Chapter 2 details the literature that informed the study. In Chapter 3, the theoretical considerations of the research are described. Chapter 4 discusses the data, methods, and materials used in this research project. In Chapter 5, the construction and validation of the 3D morphable model of the face are described. Chapter 6 describes the 3D-from-2D face reconstruction approaches developed as well as the geometric accuracy of 3D facial reconstructions using the ground-truth 3D face scans. Finally, Chapter 7 provides general conclusions and recommendations for future work.

Chapter 2. Literature Review

This chapter describes the methods for collecting facial data used in the diagnosis of FAS. Furthermore, the benefits and limitations of each data collection method are discussed. Three-dimensional (3D) surface reconstruction methods are presented with the benefits, limitations, and areas of application outlined. The chapter ends with a description of the gaps identified in the literature which informed the project scope.

2.1 Direct anthropometry

Direct anthropometry is a data collection method used to obtain measurements from body structures using physical instruments (Farkas, 1994). It uses physical instruments such as rulers or measuring tapes to obtain linear distances between defined features, protractors to obtain angles, and calipers to obtain distances between two opposite sides of an object.

For FAS facial screening purposes, the direct anthropometry method is employed to obtain facial measurements from facial characteristics related to the syndrome including palpebral fissure lengths, upper lip thinness, and philtrum flatness (Moore *et al.*, 2002). Direct anthropometry has been widely applied as an aid to understand congenital structural anomalies, where clinicians extract measurements from specific facial features using physical instruments and then perform comparison studies. The advantages of direct anthropometry are its simplicity, reliability, and affordability. Another benefit of direct anthropometry is that it allows measuring of areas of the face covered by the hair. However, this method is time-consuming when conducting mass population screening. Furthermore, participants need to comply when obtaining facial measurements; this is challenging especially children or infants since their attention span is short. Additionally, it is impossible to acquire follow-up facial surface measurements in the absence of participants. Finally, the philtrum (vertical groove between the nose and the upper lip) which is a vital pattern related to FAS facial phenotype, is rarely evaluated with this method. Several studies have used direct anthropometry to acquire data for performing comparative analysis and some are described below.

Moore *et al.* (2001) employed a direct anthropometry method, to acquire 21 craniofacial measurements from 100 individuals exposed to alcohol (41 FAS and 59 partial fetal alcohol syndrome (pFAS)) and 31 subjects not exposed to alcohol (healthy controls (HC)). The study employed multivariate discriminant analysis to classify three groups (FAS, pFAS and HC). Six of the 21 craniofacial measurements accurately differentiated between alcohol and non-alcohol exposed individuals with a sensitivity of 98%, a specificity of 90%, and an accuracy of 96%. Sensitivity is the percentage of true positives that are correctly classified as positives in the dataset; specificity is the percentage of true negatives that are correctly categorized as negatives in the dataset; and accuracy is the percentage of true results (both positive and negative) in the dataset (Moore *et al.*, 2001). Researchers differentiated between FAS and HC groups with 100% sensitivity and specificity. Additionally, the researchers indicated that 5 variables discriminated FAS, pFAS, and HC groups with an accuracy, a sensitivity, and a specificity of 88%, 86% and 94%, respectively.

Moore *et al.* (2002) obtained craniofacial measurements using direct anthropometry to quantify the face phenotype of 100 individuals exposed to alcohol and 31 healthy individuals. The researchers reported that 19 of the 21 craniofacial measurements indicated a significant difference between FAS and control groups, as well as FAS and pFAS groups. However, two craniofacial measurements were less indicative. Additionally, pattern profiles using mean z-scores found that the FAS group was significantly shorter than the pFAS and control groups. The direct anthropometry method of acquiring facial data for FAS assessment is subjective to the operator (Ort *et al.*, 2012).

In clinical and research settings, several challenges still exist for successful diagnosis of fetal alcohol spectrum disorder (FASD), which includes FAS as the most severe form. For example, obtaining a report of maternal alcohol intake depends on the birth mother's willingness to disclose information about alcohol consumption during pregnancy. Sometimes birth mothers are no longer in the life of the children at the time of examination, particularly for those children in foster care homes and adoption centres (Astley, 2006). The classic facial features of FAS tend to diminish as individuals develop into adults (Streissguth et al., 1992) although some researchers maintain that those facial features that change with age aren't specific or sensitive to prenatal alcohol exposure

(Astley et al., 2001; Chudley et al., 2007). An additional challenge is some individuals with FAS who show some, but not all, the clinical features and yet they have a history of alcohol exposure (Aase, 1994).

2.2 Indirect anthropometry

This section describes the indirect anthropometry methods for data collection which have been used for face analysis. The methods described here include two-dimensional (2D) photogrammetry, sparse 3D stereo-photogrammetry, and 3D surface imaging.

2.2.1 2D photogrammetry

Several 2D approaches for analysing FASD facial features have been discussed in the literature. These approaches employ digital cameras to collect photographs of the face for facial analysis. Image acquisition is instantaneous which minimizes motion artifacts, and images can be stored for future use. Additionally, 2D photogrammetry allows repeated measurements in the absence of a patient, which is not the case for direct anthropometry.

Multivariate statistical methods and geometric morphometrics have been used to understand the facial phenotype associated with FAS or FASD from 2D images. Geometric morphometric methods use landmark or outline information of the shape for quantitative analysis. The benefit of geometric morphometry over landmark-based anthropometry is its ability to describe the overall facial shape with few landmark measurements which are statistically unrelated (Halazonetis, 2004).

Clarren et al. (1987) employed shape information obtained from photographs to evaluate the facial effects of fetal alcohol exposure. Two methods were used to analysis the facial shapes. First, expert clinicians were asked to differentiate between alcohol exposed individuals and non-exposed individuals. Second, a morphometrics method was applied to photographs to determine facial shape features associated with fetal alcohol exposure.

Huang et al. (2005) used multidimensional discriminant analysis and principal component analysis to automatically classify images into FAS positive and FAS negative. The classifier achieved 70% accuracy.

However, 2D photogrammetric measurements are prone to magnification and distortion challenges due to varying lighting conditions and object-camera distances. Furthermore, 2D photogrammetry does not capture depth information. The shortcomings of 2D analysis have prompted the research community to explore alternative methods which leverage the 3D nature of the human face.

2.2.2 Sparse 3D Stereo-photogrammetry

Sparse landmark 3D stereo-photogrammetry method uses stereo images (multiple photographs) to obtain geometric 3D information by estimating positions of common points or landmarks. Sparse 3D stereo-photogrammetry aims to obtain 3D coordinates of corresponding landmarks from stereo images. The drawbacks of direct anthropometry and 2D photogrammetry methods are addressed using this method. The benefit of the sparse 3D stereo-photogrammetry approach is its ability to represent information in 3D. However, the stereo-photogrammetry approach has its shortcomings including occlusions and visibility constraints.

Meintjes *et al.* (2002) developed a stereo-photogrammetric tool to obtain facial data for the diagnosis of prenatal alcohol-affected children. The developed tool was used to acquire photographs of 44 subjects and stereo-photogrammetric techniques were employed to obtain data from facial features. The facial features included interpupillary distance (IPD), inner canthal distance (ICD), and palpebral fissure length (PFL). However, the technique was unable to measure other facial features like the philtrum, nasal bridge, and the upper lip.

Douglas *et al.* (2003) used the stereo-photogrammetric tool developed by Meintjes *et al.* (2002) to capture photographs for diagnosing the facial phenotype associated with FAS. They presented an algorithm that extracts and measures four facial distance features from stereo-photographs for FAS analysis. The facial features assessed were PFL, IPD, ICD, and outer canthal distance (OCD). The comparison between the proposed algorithm and manual method reported a mean absolute

difference of less than 1 mm for both PFL and IPD. The study established that the measurements for ICD and OCD did not compare well with manual measurements.

Mutsvangwa *et al.* (2007) employed a stereo-photogrammetric tool developed by Meintjes *et al.* (2002) to obtain landmarks from stereo-photographs of 34 subjects (20 normal controls & 14 FAS) aged between 6 to 7 years for the diagnosis of FAS. The researchers then used statistical techniques such as generalized Procrustes analysis (GPA), principal component analysis (PCA), and linear discriminant analysis (LDA) to analyze the facial data. Their analysis revealed significant differences in facial shape information between the two groups (normal and FAS). Additionally, Mutsvangwa *et al.* (2009) developed an improved stereo-photogrammetric tool to obtain 3D information from stereo-photographs of infants aged 5-12 years for the diagnosis of FAS.

2.2.3 3D surface imaging systems

Several 3D facial data capturing systems have been used to overcome the limitations of 2D photogrammetry and sparse landmark 3D stereo-photogrammetry methods. The 3D data capturing systems include 3D computed tomography, laser 3D scanning, structured light 3D scanning, and dense 3D surface stereo-photogrammetry (Chang *et al.*, 2015; Honrado *et al.*, 2004; Lekakis *et al.*, 2016; Tzou *et al.*, 2011). For the detailed description of these imaging modalities, the reader is directed to paper by Lekakis *et al.* (2016). These imaging systems can be used to generate 3D surface information from objects such as human faces. The choice of the imaging system is influenced by the acquisition time (capture speed), ease of use, its effect on the skin, and the areas of application. Some of the 3D surface imaging systems used to obtain facial data are discussed below.

3D computed tomography (3DCT): In 3DCT, the 3D face surfaces are acquired using the principle of surface rendering (Lekakis *et al.*, 2016). Rendering is defined as the process of creating a 3D image from an object. This modality is not often regularly used due to radiation exposure which may be hazardous to tissue and bone. Furthermore, CT has a low soft tissue resolution which may limit acquisition of surface details that are essential in face analysis.

Laser 3D scanning: Laser scanners employ the principle of surface triangulation to create 3D surfaces of objects (Largo *et al.*, 2013; Lekakis *et al.*, 2016). The scanner emits a laser pattern that sweeps around the object. A camera is used to record the data points and generate a geometrical structure of the object (Largo *et al.*, 2013). 3D laser scanners produce accurate and high-resolution images. The image acquisition process when using laser scanners is fast and requires no contact with or distortion of the object being scanned. However, laser scanners are costly and prone to producing distortions in scans of children because they are unable to hold the same pose for long. Additionally, these scanners produce laser light which may be hazardous to the eyes; subjects are required to close their eyes as a protective mechanism otherwise they stand a risk of permanent damage to the retina (Honrado *et al.*, 2004).

Structured light 3D scanning: This surface imaging method involves the projection of parallel patterns of coded or structured light on an object (Lekakis *et al.*, 2016). The scanner consists of a projector and digital cameras. Typically, a narrow beam of light is projected onto the object's surface and arbitrary fringes are acquired by the digital cameras (Lekakis *et al.*, 2016; Tzou *et al.*, 2011). The images are recorded simultaneously from different viewpoints and processed to generate 3D images. The primary benefit of structured light scanners is the high capture speeds, in microseconds. The downside of structured light scanners is high costs.

Dense 3D surface stereo-photogrammetry: In contrast to structured light 3D scanning, 3D surface stereo-photogrammetry does not need a special projection pattern. Similar to laser scanning, 3D stereo-photogrammetry works on the principle of surface triangulation where random unstructured light patterns are combined with visible natural patterns on the object's surface to generate surface geometry and texture of an object (Lekakis *et al.*, 2016). Similar to laser scanning and structured light scanning, the dense 3D surface stereo-photogrammetry method is affected by occlusions.

Moore *et al.* (2007) used, indirect anthropometry, specifically a laser scanner, to obtain facial data for face analysis. The study populations were grouped into North American Caucasian (NAC), African American (AA), Finnish Caucasian (FC) and Cape Coloured (CC). The researchers employed a discriminant analysis to perform a classification study. Eight facial variables correctly categorized the FC sample with a 93% accuracy, and five facial variables were used to classify the

CC sample with a 92% accuracy. The study used 2 variables to categorize the AA sample with an accuracy of 79%. The NAC sample was classified with 77% accuracy using inner canthal width (ICW) and outer canthal width (OCW). The most discriminant facial features across the 4 study samples were palpebral fissure length, ICW, and OCW.

Hammond *et al.* (2004) developed dense surface models (DSMs) from 3D face scans obtained using DSP400 and MU2 face scanners (http://www.3dMD.com) and then used the developed DSM to reconstruct 3D surface data for studying the facial morphology. The study employed a 10-fold cross-validation testing to discriminate the samples based on pattern recognition algorithms including nearest mean, neural networks, logistic regression, and support vector machines. Their study was focused on understanding Noonan and velo-cardio facial syndromes which affect the face regions.

The 3D surface capturing systems described above tend to be expensive to acquire and maintain. This limits their use in low-resource settings. Although the cost of some the 3D scanning systems has been declining, those on the cheaper side of the cost spectrum tend to produce low-resolution images making them unsuitable when analyzing facial images for medical purposes. The approximation of 3D facial data from 2D images of the face may offer a cheaper alternative to 3D surface capturing systems.

2.3 3D-from-2D surface reconstruction methods

Morphable models provide an approach for creating 3D surface models from 2D image data. Section 2.3.1 describes 3D morphable models (3DMMs) of human faces, which aid in 3D face reconstruction. Section 2.3.2 discusses the methods for reconstructing 3D human faces by fitting 3DMMs to single 2D facial images.

2.3.1 3D morphable models

Three-dimensional morphable models (3DMMs) are high resolution statistical models used for modelling and representing 3D human faces (Blanz *et al.*, 1999; Booth *et al.*, 2017; Booth *et al.*, 2018; Dai *et al.*, 2017). The 3DMMs are constructed from a collection of 3D face scans usually obtained with 3D surface capturing systems. The face scans represent shape and texture variations

seen in the population. The 3DMMs may be categorized as either statistical shape models (SSMs) or statistical shape and appearance models (SSAMs) (Cootes *et al.*, 2001). The SSMs contain shape variation information only, while SSAMs contain both shape and texture variability information. Some of the 3DMMs of the faces developed in the literature are described below.

Blanz *et al.* (1999) demonstrated an approach for building 3D morphable models from a set of face scans by computing correspondences with an optical flow algorithm. The shape and texture variations in a collection of face scans were encoded as vector representations, and point-to-point correspondences across the dataset were computed. This approach was used by Paysan and colleagues to construct a 3D morphable model of human faces known as Basel Face Model (BFM) (Paysan *et al.*, 2009). The BFM was built from a collection of 200 face scans. The establishment of dense correspondences across a collection of face scans was facilitated by a sparse set of manually selected facial landmarks. The non-rigid iterative closest points (NICP) method (Cheng *et al.*, 2017) was used to register a template face mesh to each face scan in the training data.

A recent study by Booth *et al.* (2018) presented a 3D morphable model of human faces known as the Large Scale Facial Model (LSFM) which was constructed from 9,663 facial scans. The study used the NICP method for registration of the template surface to target face scans, generalized Procrustes analysis (GPA) for similarity alignment of the face scans, and principal component analysis (PCA) for statistical analysis of the registered face scans to construct a 3D morphable model. The constructed model included face scans from a diverse population with varying age, gender, and ethnicity, resulting in the largest morphable face model ever constructed to date as suggested by the researchers. However, additional research focusing on human face variations would still be required before the morphable model could be used for medical purposes (Booth *et al.*, 2018).

Dai *et al.* (2017) built a 3D morphable model known as the Liverpool-York Head model (LYHM) from complete (full) head scans. The LYHM was constructed from 1200 head-face scans. The dense correspondence between the template surface and training facial scans was established by using a hierarchical parts-based template morphing method for the shapes and optical flow refinement method for the textures. The authors applied a GPA approach to remove similarity

effects in the training head scans followed by a PCA to build the 3DMM of the head-face scans. Additionally, in this research, age and gender specific 3DMMs were constructed and compared. However, in their analysis, the authors found large variation in the chin region of the face and thus more research was required to validate the approach (Dai *et al.*, 2017).

The applications of 3DMMs are varied. The shape and texture models have been used in the study of morphometrics of anatomical structures including human faces (Bacivarov *et al.*, 2009; Hammond *et al.*, 2004). Other applications of 3DMMs include segmentation of medical images, extrapolation of shapes from sparse information, and surface approximations for surgical planning and assessment (Heimann *et al.*, 2009; Rajamani *et al.*, 2007). Finally, 3DMMs have been used as aides to estimate 3D surface reconstructions from partial data or 2D images (Booth *et al.*, 2018).

2.3.2 Model-based methods for 3D-from-2D reconstruction

Several 3D-from-2D reconstruction methods based on fitting a 3DMM to a 2D image of the face have been described in the literature (Arellano *et al.*, 2012; Blanz *et al.*, 1999; Blanz *et al.*, 2003; Hu *et al.*, 2017; Romdhani *et al.*, 2002; Romdhani *et al.*, 2005; Schönborn *et al.*, 2016). Some of the 3D-from-2D reconstruction methods using 3DMM fitting processes are discussed below.

Blanz *et al.* (1999) demonstrated a fitting approach for approximating the 3D shape and texture of a human face from a single image using a 3DMM. The researchers addressed two challenges: (1) automatic registration of new faces by calculating dense correspondences, and (2) regularization of the modeled faces to preserve realism in the reconstructed faces. However, the geometric accuracy of the face reconstructions was only qualitatively assessed, and a quantitative assessment is necessary before adoption for medical applications.

Romdhani *et al.* (2002) presented a linear fitting approach for estimating a face shape and texture viewed at different poses and illumination conditions. The approach employed linear equations based on optical flow to determine shape and texture coefficients of the face. This fitting approach generated a 3D face reconstruction from a single 2D image by applying an analysis-by-synthesis technique. The fitting algorithm was evaluated considering the combined pose-illumination variations for face identification accuracy giving a success rate of 60%. However, the authors

recommended that the fitting algorithm be tested on a larger database and also that the geometry of the face reconstructions be evaluated.

Blanz *et al.* (2003) presented an automatic method for estimating the 3D face from an input 2D image by fitting a morphable model using intensity-based cost function with a stochastic newton optimization algorithm. The fitting algorithm estimated shape and texture parameters as well as pose and lighting condition parameters of the face. The dense correspondence (3D-from-2D non-rigid registration) between the reference face model and the target image of the face was established with an optic flow algorithm (Blanz *et al.*, 1999). The estimated faces were only qualitatively compared in a recognition task. The geometric accuracy of the 3D face reconstructions was not quantified.

Romdhani *et al.* (2005) proposed multi-feature fitting (MFF) methods to estimate intrinsic and extrinsic parameters of the face surface from a single input image using the Levenberg-Marquardt optimization technique (Kabus *et al.*, 2004). The method used both pixel intensities and image cues such as edges and specular highlights to approximate a 3D face surface. This resulted in a cost function that is smooth and easy to minimize (Romdhani *et al.*, 2005). The research study did not report on the geometric accuracy of the 3D face reconstructions.

Arellano *et al.* (2012) presented a mean shift method for fitting shape models to a set of observations without the need for finding correspondences. The method globally estimated a surface reconstruction from a point cloud with shape information following a Bayesian framework. The average reconstruction error of the method was 1.5 mm with a standard deviation of 2.9mm. However, the researchers used synthetic faces generated from Basel face model to reconstruct 3D faces for evaluating the reconstruction errors. It would be interesting to apply this approach on real 2D images for estimating 3D surface reconstructions for medical applications.

Hu *et al.* (2017) proposed a fitting method - called efficient stepwise optimization - to reconstruct a 3D face from a single 2D image using a 3DMM. This fitting process optimizes the shape, pose, light strength, light direction and texture parameters, separately to generate a 3D face reconstruction from a single 2D image. The shape reconstruction performance computed by cosine

similarity was 0.5. The cosine similarity is a metric used to measure the similarity between surfaces by considering the vector angles; it measures surface orientations but not magnitudes.

Schönborn *et al.* (2016) proposed a probabilistic method for estimating a 3D shape and texture of the face by fitting a Gaussian process morphable model to a single image. The method employed a Bayesian framework to infer a posterior distribution model conditioned on the set of the observations. This method was based on the Metropolis-Hastings algorithm making it a sampling-based fitting algorithm. The advantage of this method was that there was no need to compute gradients making it less prone to local minima compared to other standard fitting methods. The face reconstruction performances were computed based on 2 databases (BFM & BU-3DFE) giving root mean square averages of 3.79 mm and 5.39 mm, respectively.

2.4 Summary of the literature review

The traditional method for obtaining the facial data for assessment of FAS is direct anthropometry. This method is simple to apply, reliable, and affordable, yet time-consuming and intrusive. The 2D photogrammetry of the indirect anthropometry approaches, provides advantages as compared to direct anthropometry but does not capture depth information, which leads to inadequacies when analyzing the face. 3D surface imaging systems have the advantage of providing the full facial geometry allowing for a comprehensive analysis of the face. However, the advantages of 3D surface imaging systems come at a cost which limits their use in resource limited settings.

The approximation of 3D facial surface from single 2D images, for example with the use of 3DMMs may offer a cheaper alternative to 3D surface capturing systems. For human face applications, the reconstruction is performed by fitting a 3DMM onto the target facial image generating a 3D face reconstruction. However, there is limited reporting in the literature on the geometric accuracy of 3D face reconstructions; this has implications for the use of such methods in medical applications. Additionally, to the best of the author's knowledge at the time of writing, model-based 3D-from-2D reconstruction of the face has not been reported for FAS applications.

Chapter 3. Theoretical Considerations

This chapter introduces the theory and methods related to statistical shape and appearance models of three-dimensional (3D) human face surfaces. Section 3.1 contains a brief description of the anatomy of a human face, introduces in-silico human face surface representations, and provides definitions for relevant facial landmarks. Section 3.2 details methods for evaluating the landmarking quality. Section 3.3 describes the correspondence problem in the modeling process. In section 3.4, the registration methods used to solve the correspondence problem are outlined. Section 3.5 focusses on Gaussian processes and how they are used to develop Gaussian process morphable models (GPMMs). In section 3.6, the metrics used to evaluate model quality are explained. Section 3.7 presents the methods used to fit a statistical model to an image. The metrics used to evaluate 3D surfaces are discussed in section 3.8. Finally, section 3.9 summarizes the chapter.

3.1 Representation of the human facial surface

The structure of the bones underlying the soft tissues in the neurocranium determines the human facial appearance (Sykes *et al.*, 2018). The facial skeleton consists of the frontal bone, superiorly; the bones of the midface; and the mandible, inferiorly. The facial bones provide structural support and protect facial organs such as ears, nose, and eyes. Additionally, facial bones serve as attachment sites for muscles controlling facial expressions and mastication (Bentsianov *et al.*, 2004).

The facial regions are made up of features that are unique due to anatomical variations in the neurocranium. The differences observed in the human face are related to the outline of the eyebrows, contours of the nose, the borders of the mouth, the prominence of the cheeks, skin texture properties, and spacing of the eyes. Other variations in the human face are attributed to extrinsic factors such as pose, facial expressions, and illumination conditions. These facial variations are used to differentiate human faces in fields such as computer vision and for biomedical applications. By interpreting and understanding these facial patterns in a given population, several features may be leveraged to detect and diagnose certain syndromes that affect the human face.

A 3D face surface or surface mesh typically consists of shape and texture information. The geometry of a 3D surface scan is represented by a set of vertices (points with coordinates) and a list of triangles joining these vertices. The vertices consist of ordered triples of X, Y and Z coordinates for each i^{th} vertex as a representation of its position in 3D space. The combination of vertices and triangles describes a surface mesh with a dense cloud of points in space. The vertices of a surface mesh are also known as points or landmarks. Facial landmarks serve as anchor points for registration methods. The texture of the facial surface is associated with the surface mesh as a per-vertex colour vector consisting of texture coordinates. The texture coordinates are represented by RGB pixel values on each vertex.

Landmarks are defined as meaningful points which represent surface meshes with either 3D or 2D coordinates (Aubert *et al.*, 2016). Facial landmark annotation is the process of detecting and localizing landmarks on the face surface (Celiktutan *et al.*, 2013). The annotation process is an important step for registration and reconstruction tasks in statistical shape model analysis. The process of selecting landmarks is extremely challenging due to variability in 3D facial surfaces and other extrinsic factors including lighting conditions, pose, facial expressions, and facial occlusions. Landmark-based methods use less computing resources and allow for faster algorithm execution.

Facial landmarks can be categorized into primary (fiducial) and secondary (ancillary). The primary facial landmarks can be identified more reliably and include the corners of the eyes, the nostril corners, the mouth corners, the endpoints of the eyebrow arcs, the ear lobes, and the nose tip. Secondary facial landmarks are typically determined through interpolating primary landmarks. The secondary landmarks include surface landmarks on the nostrils, chin, nasion, eyelids, cheek contours, and eyebrow midpoints.

3.2 Evaluation of landmarking quality

Facial landmarking is the process of identifying and selecting key points on the facial surface. The quality of landmarking can be assessed by evaluating observer precision and reliability.

Reliability is the degree to which measurements are consistent. One of the statistics used to evaluate observer reliability is the intraclass correlation coefficient (ICC) (Cook, 2000). The ICC values range from 0 to 1 with values approaching the upper limit of 1.0 indicating a perfect agreement in landmark selections, whereas the ICC values tending towards the lower limit of 0 indicate an increasing disagreement in the landmark selections. An ICC value greater than 0.75 is considered excellent (Victor *et al.*, 2009).

Precision is defined as the measure of reproducibility of measurements. The mean absolute difference (MAD) may be used to estimate observer error of repeated landmark measurements of the same quantity (Aldridge *et al.*, 2005; Mutsvangwa *et al.*, 2011). The MAD is defined as the average absolute differences between repeated measurements obtained across the dataset.

The intra-observer precision is calculated by comparing the two sets of 2D landmark measurements obtained by the same observer. The expression for computing intra-observer precision between two sets of 2D landmark measurements is shown in equation (3.1).

$$intraobserver\ 2D\ error = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{3.1}$$

where n is a number of sets of landmark measurements, coordinates x_i and y_i represent the observations of the i^{th} landmark measurements and $\bar{x}$ and $\bar{y}$ represent the mean position of landmark coordinates across all the recordings by the same observer.

The formulation used to compute intra-observer precision between two sets of 3D landmark measurements is shown in equation (3.2) (Mutsvangwa *et al.*, 2011).

$$intraobserver\ 3D\ error = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2} \tag{3.2}$$

where n is a number of sets of landmark measurements, coordinates x_i, y_i, and z_i represent the observations of the i^{th} landmark measurements, and $\bar{x}$, $\bar{y}$, and $\bar{z}$ represent the mean position of landmark coordinates across all the recordings acquired by the same observer.

The inter-observer precision is calculated by comparing average 2D landmark measurements acquired by two independent observers, as shown in equation (3.3).

$$interobserver\ 2D\ error = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(\bar{x}_{1i} - \bar{x}_{2i})^2 + (\bar{y}_{1i} - \bar{y}_{2i})^2} \quad (3.3)$$

where n is a number of sets of landmark measurements, coordinates $\bar{x}_{1i}$ and $\bar{y}_{1i}$ are the averages of the landmark measurements obtained by different observers, and the coordinates $\bar{x}_{2i}$ and $\bar{y}_{2i}$ represent the grand mean of the observations acquired across all the different observers.

The inter-observer precision for 3D landmarks is calculated using equation (3.4) (Mutsvangwa *et al.*, 2011).

$$interobserver\ 3D\ error = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(\bar{x}_{1i} - \bar{x}_{2i})^2 + (\bar{y}_{1i} - \bar{y}_{2i})^2 + (\bar{z}_{1i} - \bar{z}_{2i})^2} \quad (3.4)$$

where n is a number of sets of landmark measurements, coordinates $\bar{x}_{1i}$, $\bar{y}_{1i}$, and $\bar{z}_{1i}$ are the averages of the landmark measurements obtained by different observers, and the coordinates $\bar{x}_{2i}$, $\bar{y}_{2i}$, and $\bar{z}_{2i}$ represent the grand mean of the observations acquired across all the different observers.

The threshold values for interpreting precision estimates proposed by Aldridge *et al.* (2005) and Mutsvangwa *et al.* (2011) were used in this study, as shown in Table 3.1.

Table 3.1: Error ranges indicating acceptable precision levels

Error intervals	**Precision levels**
Less than 1 mm (0 - 0.9 mm)	High precision
Between 1 mm - 1.5 mm	Precise
Between 1.6 mm - 2.0 mm	Moderate precision
Greater than 2.0 mm	Imprecise

3.3 Correspondence problem

The correspondence problem is the problem of finding a set of landmarks in one object that matches another set of landmarks in a second object of the same object class (van Kaick *et al.*, 2011). Although the objects may be of the same class, they can be captured from different points

of view or acquired at different times, which changes the global coordinate system of the objects. The goal of correspondence establishment is to find meaningful relationships between corresponding sets of features in objects. The correspondence problem is easier to state but hard to define mathematically. The approaches for establishing correspondences across objects may either be manual or automatic. The manual approach involves an expert annotating surfaces consistently with a set of landmarks. This approach is time-consuming, error-prone, and impractical when dealing with large amounts of data, or when trying to establish dense correspondence across surfaces. The automatic approach employs algorithms that compute correspondences across a set of examples in the dataset. This approach is fast and convenient when processing large amounts of data and is the feasible approach for establishing dense correspondence. For instance, given two surfaces, the correspondence is established for every point on one surface referred to as reference, by finding the matching point on another surface denoted as the target. Typically, a registration procedure is employed in the automated approaches to establishing correspondences across the dataset.

3.4 Registration

Registration is the task of computing correspondences that bring a set of selected points on a reference surface to a set of corresponding points on the target surface of same object class (Oliveira *et al.*, 2014; Viergever *et al.*, 2016). Two classes of registration exist; rigid and non-rigid. In rigid registration, surfaces are registered by applying rigid transformations. For non-rigid registration, surfaces are registered by applying surface deformation techniques between the reference surface and the target surfaces.

3.4.1 Rigid alignment

A common method for performing a rigid alignment of surfaces is the iterative closest point (ICP) algorithm (Besl *et al.*, 1992; Rusinkiewicz *et al.*, 2001). The ICP algorithm aligns two surfaces by minimizing the difference between a set of landmarks on one surface and a closely corresponding set of landmarks on the second surface. Once the landmarks have been identified and the relationships between the corresponding landmarks have been established, a rigid transformation that brings the reference landmarks on one surface closer to the fixed set of landmarks on another

surface is computed. The reference landmarks are transformed to the fixed landmarks based on the optimal values of the rigid transformation parameters. The application of the computed rigid transformation parameters results in a new surface that is aligned to the reference surface. Several researchers have looked into developing accurate rigid alignment methods (Champleboux *et al.*, 1992; Szeliski *et al.*, 1994). Although rigid alignment methods show a good estimation when registering simple anatomical structures, most structures inside or outside the human body are complex in nature. The rigid alignment methods struggle when computing correspondences between complex anatomical structures such as human faces.

3.4.2 Non-rigid surface registration

Non-rigid registration is a task of computing deformations that map a set of elements on the reference surface onto a set of elements on the target surface in the same coordinate system (Oliveira *et al.*, 2014; van Kaick *et al.*, 2011). The goal of non-rigid registration is to recover the deformations of the reference surface that moves it close to the target surface. The computed deformations affect the geometrical attributes of the surfaces. Variations in surface geometry may be caused by factors such as noise or the elastic nature of anatomical objects. Given elastic objects such as human faces, non-rigid registration methods are preferred when performing a full surface registration since these methods give more degrees of freedom on how these surfaces can be registered (van Kaick *et al.*, 2011).

A common method used for non-rigid registration of surfaces is the non-rigid iterative closest point (NICP) algorithm (Allen *et al.*, 2003; Amberg *et al.*, 2007). This registration method employs the closest point technique to determine deformations that deform a reference surface to match each target surface. The steps for implementing the NICP algorithm as described by Amberg *et al.* (2007)

Firstly, given a collection of points on a reference surface, determine the closest set of points on a target surface by computing point-to-point correspondences between the reference and target surfaces. The point-to-point correspondence is achieved by searching for every point on the target surface and matching it to the closest point on the reference surface. Once the two surfaces are in partial correspondence, an affine transformation that deforms the reference surface to match the

target surface is computed. This affine transformation is a combination of the rotation, translation, and scaling parameters. The differences between the deformed reference surface and target surface are optimized using a distance metric. The strong deformations in the surfaces are regulated by enforcing a regularization term. The outcome for applying affine transformations is a deformed surface which is as close as possible to a target surface. Once the two surfaces are locally aligned, the iterative closest point technique is applied to bring the two surfaces in dense correspondence. The deformed surface is said to in dense correspondence if the surfaces have the same number of points in similar anatomical locations.

Another method used in non-rigid registration of surfaces is the B-splines approach (Huang *et al.*, 2006). This approach models the local deformations of an object embedded in space by manipulating the underlying surface mesh. This approach computes free-form deformation models (FFDMs) by modifying control points of an object through changing the bounding box.

A recent state-of-the-art non-rigid registration method is the Gaussian process (see section 3.5 for a description of Gaussian processes) fitting approach (Gerig *et al.*, 2014; Lüthi *et al.*, 2013). This approach is based on a morphable model fitting to the target surface. The Gaussian process morphable model (GPMM) (Lüthi *et al.*, 2017) computes deformation fields that compares the model to a target surface, and the best deformations are selected to represent the target surface. Here, the registration method is formulated as the minimization problem that estimates the error function by computing the difference between the deformed surface and the target surface.

3.4.3 Similarity Alignment

After registration of face surfaces, these surfaces may still be misaligned. The misalignments may be removed by performing rotation, translation, and rescaling. Given two surfaces, a reference surface and a target surface with their associated landmarks, the aim is to compute a spatial transformation that maps the reference landmarks on the surface to its corresponding landmarks on the target surface. One of the methods used to perform a similarity alignment is the generalized Procrustes analysis (GPA) (Gower, 1975). The GPA approach iteratively aligns three or more

surfaces, each time recalculating a mean and then repeating the process until there is no longer a change in the mean surface.

3.5 Gaussian process morphable models

A Gaussian process (GP) is defined as a set of variables with a joint Gaussian distribution over all possible functions (Rasmussen, 2004). The GP is a type of a stochastic process which describes a multivariate probability distribution over functions. The GP provides a continuous representation defined by mean, μ and covariance function, k as expressed in equation (3.5).

$$GP(\mu, k) \tag{3.5}$$

Gaussian processes use the marginalization property to model discrete representations from continuous Gaussian representations. The marginalization property is achieved by restricting the distribution to a finite set of points on a surface. In surface modelling, marginal distributions can be used to compute the confidence regions of a surface mesh.

A statistical model is constructed from a collection of 3D surface scans where correspondences have been established. For a statistical model based on a GP, the mean function represents a deformation field that deforms a reference surface using a mean surface. The covariance function defines the likely deformations in the statistical model. The covariance function is also known as the kernel. This kernel is asymmetric and positive semi-definite function (Lüthi *et al.*, 2017). The surface deformations, u, from the GP distribution are sampled to generate new deformations by interpolating the reference surface. Additionally, a GP is represented as an orthogonal set of basis functions using the Karhunen-Loeve expansion (Berlinet *et al.*, 2011). The new surface deformation is estimated as linear combinations of surfaces using the expression in equation (3.6).

$$u = \mu + \sum_{i=1}^{\infty} \beta_i \sqrt{\lambda_i} \emptyset_i, \quad \beta_i \sim N(0, 1) \tag{3.6}$$

where λ_i is an eigenvalue, ϕ_i is eigenvector, μ is the mean surface, β_i is a set of parameters of the statistical model drawn from a Gaussian process distribution, and u denotes a new surface deformation estimated from the mean surface and random surface instances.

A low-rank approximation of the GP is calculated using the Nystrom (or quadrature) method (Lüthi *et al.*, 2017; Rasmussen, 2004). This method reduces the computational complexity of the training data. Once the model is constructed, the quality of construction can be validated. The next section (3.6) discusses the common evaluation metrics employed to assess the performance of the statistical model.

3.6 Evaluation of the statistical models

There are three common metrics for evaluating the quality of a statistical model, namely generalization, specificity, and compactness. Each evaluation metric measures a different aspect of the quality of the statistical model. The evaluation metrics are widely used when assessing mesh-based statistical shape models. These evaluation metrics indicate how well the model construction process transforms the original data from high dimensional space into a lower dimensional space. The evaluation metrics are described below.

3.6.1 Generalization

The most popular approach for evaluating the performance of a statistical model when generating unseen examples of itself is the leave-one-out reconstruction (Styner *et al.*, 2003). For instance, given a set of n training instances, a model can be constructed with $n - 1$ training instances. The constructed model is then fitted to the left-out instance. The procedure is repeated for all instances in the training set. The generalization error of the model is then determined by computing the surface distances between the original left-out instance and its corresponding reconstructed instance. The results of the application of the leave-one-out method indicate how well the model fits out-of-training data. Large reconstruction errors indicate that more training instances are required for better generalization of the statistical model.

3.6.2 Specificity

Specificity is the ability of a statistical model to randomly generate instances that are similar to those present in the real dataset (Styner *et al.*, 2003). Specificity is evaluated by generating random instances from a statistical model and then comparing them to the nearest instances in the dataset.

The specificity error is computed as the average distance between randomly generated instances from the statistical model and nearest instances in the dataset.

3.6.3 Compactness

Compactness is defined as a number of parameters needed to express a fraction of the variance in a collection of training data. To be regarded compact, a statistical model should have little variability and require few parameters for defining an instance (Styner *et al.*, 2003). Compactness is determined by computing the retained cumulative variance.

3.7 Methods for fitting a 3D statistical model to a 2D image

Fitting a statistical model to an image involves estimating the model parameters that map input image features to the reference model features (Mani *et al.*, 2013). Typically, this requires a model that encodes both shape and textural variation. When fitting a statistical model to an image, the goal is to obtain an instance of the model that aligns the rendered image onto the target input image, as closely as possible. To obtain a good reconstruction, the model and imaging parameters are always optimized using the fitting algorithm. The major challenge of statistical model fitting methods is determining the correspondences between the target image features and the reference model features. Additionally, most model-based fitting approaches are affected by external factors in the target image such as occlusions and facial hairs. In model fitting, the correspondence problem is generally solved by applying registration methods and determining a good set of initial model parameters. The initialization approach reduces the solution search space, search time, and associated computational resources. One of the initialization techniques is to provide a set of manually landmarks for which correspondence has been determined manually, between the reference model and the target image. The statistical model fitting process can be separated into three steps: (1) projection of the statistical model into the image plane, (2) alignment of the statistical model to the input image, and (3) fitting the statistical model with an image to estimate the parameters that relate the model to the image.

3.7.1 3D projective transformations

Three-dimensional projective transformation defines correspondences between features of a 3D object (or a 3D statistical model) and a 2D image by computing parametric transformations which map a set of 3D point coordinates in the object onto a set of 2D point coordinates in an image. Projective transformations are categorized as either perspective or orthographic (parallel) projections. In orthographic projections, the center of projection is infinitely far from the view plane, the projectors (rays) are parallel to each other, and there is always a need to define the direction of projection. For perspective projections, the center of projection is finitely far from view plane, projectors are not parallel to each other and there is no need to define the location of the center of projection.

3.7.2 Spatial transformations in model fitting

Given a model and an image, an alignment brings the statistical model as close as possible to input image data. A spatial transformation establishes correspondences between a set of points in a model and its corresponding set of points in an image. Depending on the application, the transformations can be rigid or similarity transformations. The rigid transformation includes a translation vector and a rotation matrix, whereas the similarity transformation, in addition to a translation vector and a rotation matrix, includes a scaling factor. Once the relationship between the statistical model and the image data is established, a distance metric is defined, which is the difference between the model points and the image data points. This distance metric minimizes the model parameters so that the model instance matches the given image data as closely as possible.

3.7.3 Posterior distribution

Bayesian inference is a method used to update beliefs about the statistical model given the set of observations (image data). This can be leveraged to provide an alternative way to fit 3D statistical models to images. The Bayesian approach considers statistical models as prior models during the fitting procedure; the fitting results in the generation of a posterior distribution model, which is updated after observing the image. More concretely, given the image data, I the posterior distribution model can be estimated using Bayes' theorem defined in equation (3.7).

$$P(\theta/I) = \frac{P(I/\theta)\, P(\theta)}{P(I)} \tag{3.7}$$

In equation (3.7), θ is a vector of unseen model parameters, I is a vector of observed data, $P(\theta/I)$ is a posterior distribution model, $P(I/\theta)$ is a likelihood function, $P(\theta)$ is a prior model, and $P(I)$ is a normalization factor.

In the Bayesian inference framework, fitting algorithms can be used to estimate both certain and uncertain observations provided that the likelihood is available. The likelihood function estimates the similarity in terms of distance measures between the features in the prior model and the features in a set of observations. The prior distribution model is the knowledge about the model before measuring any observations. Both the likelihood and prior model describe the posterior model after application of a normalization factor. The beliefs about the posterior distribution model are updated with each observation. The posterior distribution model may be interpreted as conditioning the prior distribution on the set of observations. The generated posterior distribution model can be employed for analysis by drawing samples using known estimation methods. The normalization factor can be intractable to compute and thus methods exist that use proportionality as mathematically expressed in equation (3.8).

$$P(\theta/I) \propto P(I/\theta)P(\theta) \tag{3.8}$$

The commonly used method is the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis *et al.*, 1953). The MH algorithm is a Markov Chain Monte Carlo method, which generates a set of random samples from a posterior distribution (Schönborn *et al.*, 2016). The random samples are then used to estimate the distribution of interest or to calculate the expected values. The MH algorithm works on the intuition that as more samples are generated, the closer their distribution approximates the target distribution. These samples are produced iteratively, where the distribution of the next sample depends on the current sample value. The sample can be accepted or rejected based on some probability criterion. The benefits of the MH method are that it can be used without gradient information, it is less prone to local optima and it incorporates information about the certainty of a fit; the main disadvantage of the MH algorithm is that it can be affected by highly correlated target distributions (Schönborn *et al.*, 2016).

3.8 Evaluation metrics for 3D surface meshes

Several evaluation metrics have been studied in the literature for faster and stable convergence of fitting methods and comparison of 3D surfaces. These evaluation metrics include Euclidean distance (Danielsson, 1980; Wang *et al.*, 2005), Hausdorff distance (Takacs, 1998), average distance (Burgstaller *et al.*, 2009), and root mean square error (Chai *et al.*, 2014).

3.8.1 Euclidean distance

This is distance between a set of points in Euclidean space (Wang *et al.*, 2005). The purpose of obtaining Euclidean distances between pairs of landmark points is to measure and compare morphometric features of objects in Euclidean space. For example, given a pair of landmarks located on either side of the eye, the Euclidean distance is obtained by computing the difference between the point positions giving the horizontal distance (length) of the eye. From a set of points, X with 3D coordinates at position (x_1, x_2, x_3) and Y with 3D coordinates located at (y_1, y_2, y_3), the Euclidean distance $d_E(X, Y)$ can be formulated as shown in equation (3.9).

$$d_E^2(X,Y) = \sum (x_i - y_i)^2$$
$$d_E(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \quad (3.9)$$

3.8.2 Hausdorff distance

This is a surface comparison metric that quantifies the proximity of two corresponding sets of points on surface meshes (Huttenlocher *et al.*, 1993). In 3D surface mesh analysis, the purpose of the Hausdorff distance is to measure the maximum distance between two sets of points on surface meshes which have been aligned. For instance, given a corresponding set of points M with 3D coordinates (m_1, m_2, m_3) and N with coordinates (n_1, n_2, n_3) on two surface meshes, the Hausdorff distance between points M and N is defined following equation (3.10).

$$d(M,N) = \max(d(M,N), d(N,M)) \quad (3.10)$$

where $d(M,N) = \max \min ||((m_1 - n_1) + (m_2 - n_2) + (m_3 - n_3))||$

In the equation (3.10), $||.||$ is the norm over the points M and N, $d(M,N)$ is the Hausdorff distance.

An improved Hausdorff distance called the modified Hausdorff distance (MHD) was proposed by (Dubuisson *et al.*, 1994). The MHD is used in the image analysis community to measure the difference between a pair of images. This metric may also be applied to quantity surfaces. The MHD is defined mathematically as in equation (3.11).

$$d(M,N) = \frac{1}{N_m}\sum ||m-n|| \tag{3.11}$$

where N_m is the number of points in the set M, and m and n are the points in sets M and N.

The MHD measures the difference between sets m and n and gives the maximum distance. The measures for Hausdorff distance are not symmetric, so the ordering of surface meshes (target, reference) must be correctly followed.

3.8.3 Average distance

The average distance is defined as the expected Euclidean distance between two points in the X convex subset of n-dimensional space (Burgstaller *et al.*, 2009). The average distance is represented as mean absolute distance or as integral with λ denoting n-dimensional Lebesgue measure. Given two points x and y, the average distance is formulated as shown in equation (3.12).

$$avgdist(X) := \mathrm{E}(\|x-y\|| = \frac{1}{\lambda(X)^2}\int_X\int_X \|x-y\| d\lambda(x) d\lambda(y) \tag{3.12}$$

3.8.4 Root mean square error

The root mean square error (RMSE) or the root mean square deviation (RMSD) is a measure of the difference between predicted values and original values (Chai *et al.*, 2014). It can be used to compute the distance between points on surfaces. Given two points P_i and Q_i the RMSE is calculated as shown in equation (3.13).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - Q_i)^2} \tag{3.13}$$

where n is the number of points in the dataset.

3.9 Summary of theoretical considerations

This chapter introduced 3D face surfaces and their representations in vector space. In face modelling, 3D face surfaces are represented using both the geometry and texture information. Registration approaches for 3D surfaces were then described. These approaches are categorized into rigid, where a global alignment of surfaces is performed and non-rigid, which registers surfaces at both the global and local levels. A key factor in surface registration is establishing correspondences across a set of 3D surfaces. The correspondence establishment can be a manual or automatic. Manual correspondence involves a skilled annotator consistently selecting landmarks on the given surfaces. Automatic correspondence uses algorithms that compute corresponding features across a set of surfaces in the dataset. Once the effective correspondence of the surfaces is completed, the surfaces can be statistically analyzed to build surface models. Methods for fitting a statistical model to an observed image were presented. The fitting approach estimates the parameters that relate the synthesized image to an input image. Finally, evaluation metrics to compute the accuracy of the 3D surface reconstructions were described.

Chapter 4. Data, Methods and Materials

This chapter details the datasets as well as the software tools used in the project. Section 4.1 describes the complete facial dataset. Sections 4.2, 4.3, and 4.4 outline the composition of the training, validation, and testing datasets, respectively. Section 4.5 demonstrates facial landmarks used in the project. Section 4.6 explains the development tools used to acquire, process, build, and analyze the facial data. Finally, section 4.7 shows a schematic overview of the methodology followed in the subsequent chapters.

4.1 Facial dataset

A publicly available database of 3D face scans known as BU-3DFE, collected by researchers from the State University of New York with ethical approval from that institution, was used in this study. The BU-3DFE database consists of high resolution 3D face scans imaged using the 3dMD face system (http://www.3dmd.com/static-3dmd_systems/) (Yin *et al.*, 2006). Scans of a total of 100 subjects (56 females and 44 males) with ages ranging from 18 to 70 years are available in the database. Each face scan contains 83 landmark annotations, 3D surface geometry and an associated surface texture captured at two angles (-45^0 and $+45^0$). The resolution of the texture image is 1300 x 900 pixels. The vertices in 3D surface mesh range from 20,000 to 35,000 polygons. Each subject in the database was scanned with 7 facial expressions (“neutral”, “anger”, “disgust”, “fear”, “happy”, “sad”, and “surprise”) (Yin *et al.*, 2006). This research study used face data with neutral expressions only. The BU-3DFE database was chosen because it includes subjects from diverse ancestral groups including 51 Caucasians, 23 East-Asians, 9 African Americans, 8 Latino-Hispanics, 5 Indian Americans and 2 Middle East Asians. A sample face from the BU-3DFE database is shown in Figure 4.1.

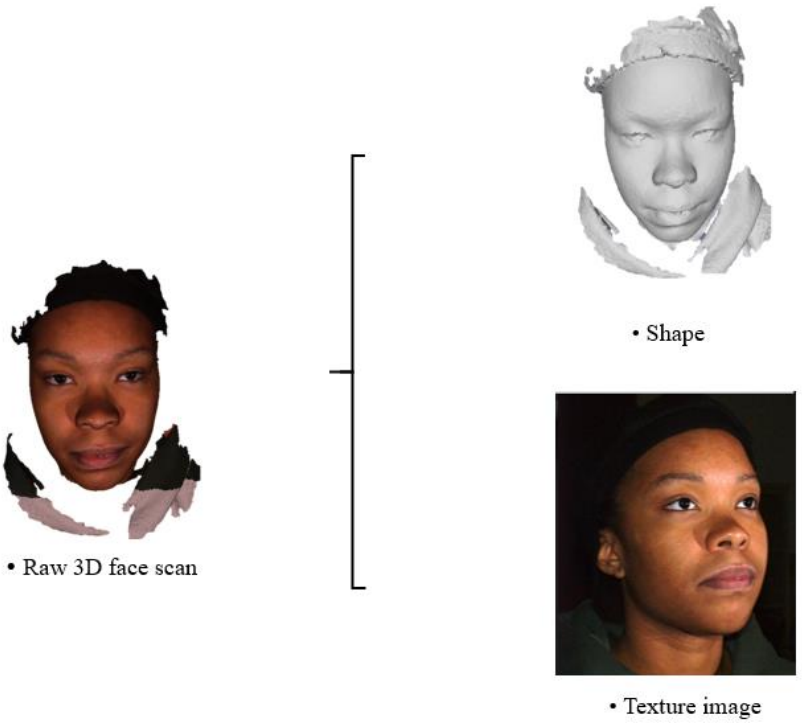


Figure 4.1: A face scan with its associated shape and a texture image.

4.2 Training data from BU-3DFE database

The BU-3DFE data was split into training and validation sets as detailed in section 4.2 and section 4.3, respectively. The training data consisted of 90 face surfaces (51 females and 39 males). All the demographic groups were represented in the training data. The training data were employed to construct a 3D morphable face model as discussed in Chapter 5.

4.3 Validation data from BU-3DFE database

A subset of the BU-3DFE dataset was used as a validation set in the project. The validation dataset consisted of ten 3D face surfaces (5 females and 5 males). The 10 selected 3D face surfaces included both biological sexes and diverse ancestral groups (5 Caucasians, 3 African Americans, and 2 Middle east Asians). The validation datasets were applied as the ground-truth face surfaces to evaluate the geometric accuracy of 3D face reconstructions from single 2D images discussed in section 6.3.

4.4 Image data

The 2D image data were extracted from the 3D face surfaces in the validation set using MeshLab software tool (see section 4.6). The image data were used as 2D target data for the reconstruction process. The facial images were resized to 512 x 512 pixels using IrfanView software tool (the tool is freely available for use on: https://www.irfanview.com/). The images were transferred into Scalismo Lab (see section 4.6 for a description) to define and select the landmark positions. Scalismo Lab is a platform for transforming and manipulating surface meshes. Twelve anatomical landmarks were selected on each facial image and a data file containing landmark information was generated. The generated 2D facial landmarks were used to guide the 3D-from-2D reconstruction process.

4.5 Facial landmarks

Mutsvangwa *et al.* (2009) identified 26 facial landmarks that are relevant for facial analysis applications related to FAS diagnosis. A subset of 12 facial landmarks were identified for this study based on ease and repeatability of selection. Table 4.1 lists 12 facial landmarks used and Figure 4.2 shows the locations of the facial landmarks on the face surface.

Table 4.1: Facial landmarks used in analysis of the face

Index	**Facial Landmarks**	**Abbreviation**
1	Right Exocanthion	exR
2	Right Endocanthion	enR
3	Left Endocanthion	enL
4	Left Exocanthion	exL
5	Right Alare	alR
6	Pronasale	pn
7	Left Alare	alL
8	Right Oral Commissure	ocR
9	Labrale Superius	ls
10	Left Oral Commissure	ocL
11	Labrale Inferius	li
12	Pogonion	p

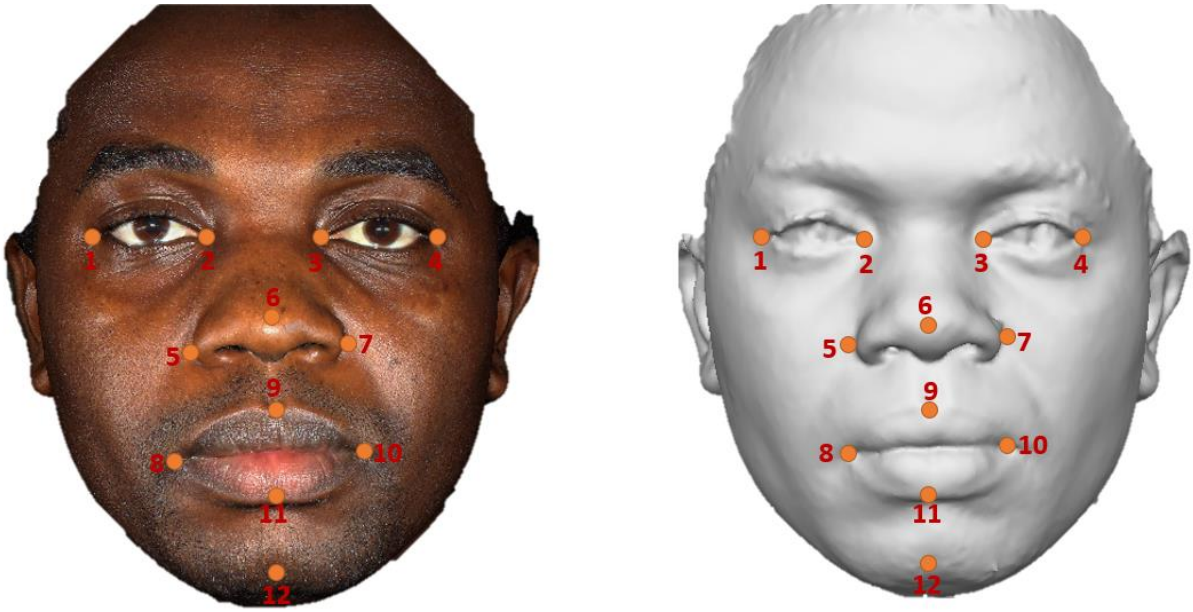


Figure 4.2: Three-dimensional face shape and texture surfaces annotated with landmarks. The dots in brown are the facial landmarks and they are indexed with numbers that are explained in Figure 4.1.

4.6 Development tools

The platforms and software used in the research project are discussed below.

IntelliJ IDEA: This is an integrated development environment (IDE) written in Java to develop computer software tools. IntelliJ IDEA is developed by JetBrains and is available in community and commercial editions (https://www.jetbrains.com/idea/). It supports many programming languages including Python and Scala.

Scalismo (Scalable Image Analysis and Shape Modelling): This is an open source tool for statistical modelling and model-based image analysis. Scalismo runs in an integrated development environment. Scalismo is developed in the Scala programming language (https://www.scala-lang.org/). Scala is a scalable programming language based on both functional and object-oriented languages. Scalismo software library is developed and maintained by the Graphics and Vision Research Group at the University of Basel, Switzerland (http://github.com/unibas-gravis/scalismo). The capabilities of Scalismo framework include building free-form deformation

models and statistical models as well as fitting the statistical models to new image data for analysis purposes.

Scalismo Lab: This is an open source tool with an executable Java jar. It runs independently of the IDE. All Scalismo functionalities are provided in one executable for small projects.

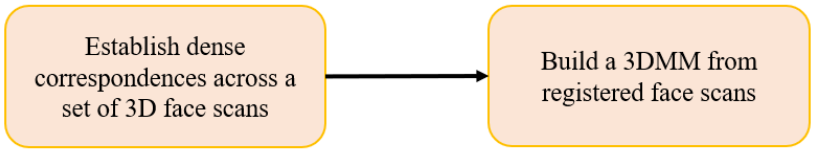
MeshLab: This is an open source platform used in processing and editing surface meshes. The functionalities of MeshLab include editing, inspecting, rendering, cleaning, converting, filtering, and texturing of mesh surfaces. MeshLab was developed by the Information Science and Technology Institute (ISTI) (Cignoni *et al.*, 2008).

Amira software: Amira (version 6.2.0) is a commercial software platform used for visualizing, processing and analyzing 3D or 4D datasets. Furthermore, Amira is applied in image segmentation, geometry reconstruction, and image registration. It was developed by Thermo Fisher Scientific and Zuse Institute Berlin (Visage Imaging: http://www.vsg3d.com).

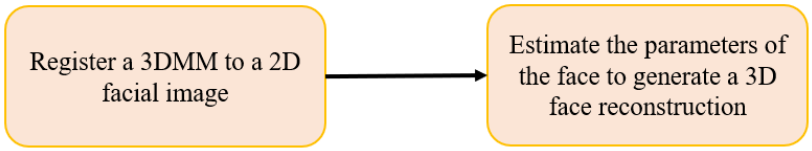
4.7 Overview of research project objectives

The project objectives in Figure 4.3 next, aimed at developing a framework for estimating a 3D reconstruction of the human face from a single 2D facial image and a 3D morphable model. An intended application (not in the scope of this study) for the 3D from 2D fitting is a tool to identify the characteristic facial features of fetal alcohol syndrome. As such, emphasis was placed on FAS related facial features in the evaluation of the framework. The project aim was decomposed into three objectives as outlined in section 1.2 and the steps associated with each objective are show in the schematic below (see Figure 4.3). Subsequent chapters describe each of the objectives.

Objective 1: Develop and evaluate the performance of a 3D morphable model of the human face



Objective 2: Estimate a 3D reconstruction of the human face from a single 2D image of a face using a 3D morphable model



Objective 3: Evaluate the geometric accuracy of 3D face reconstructions considering the facial features related to FAS

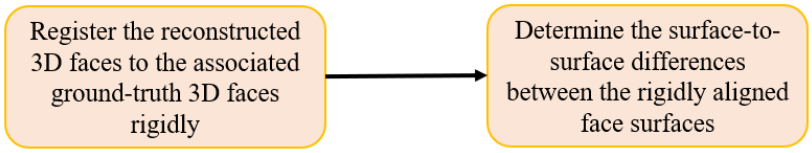


Figure 4.3: Pictorial representation of the research project objectives.

Chapter 5. Developing and evaluating the performance of a 3D morphable model

This chapter develops and evaluates a 3D morphable model (3DMM) of the face from 3D face scans. Additionally, assessment of the precision and reliability of selecting facial landmarks (used to anchor the reconstruction/registration) is reported on. Conceptually, the first step was to establish correspondences across a collection of 3D face scans in the training set. The second step was to build a 3DMM of human faces from registered face scans by performing statistical analysis on the in-correspondence training data.

The landmarking reliability and precision assessments are reported in section 5.1. In section 5.2, the preprocessing of 3D face scans is outlined. In section 5.3, the rigid registration of the 3D face scans is detailed. Section 5.4 describes the non-rigid registration of the 3D face scans using Gaussian process model fitting to achieve correspondence. Section 5.5 explains the statistical modelling of the registered 3D face scans. Samples of shape and texture instances are presented in section 5.6. In section 5.7, the performance of the 3DMM of the face is evaluated. Section 5.8 concludes the chapter with discussions on 3DMM building and performance.

5.1 Evaluation of 3D landmarking quality

Two observers participated in this part of the study. Each observer was required to select a set of 12 landmarks (see section 4.5) on each 3D face surface. The landmark selection procedure was performed twice on each face surface. A period of 24 hours was allowed to pass between the landmarking process for the two trials per observer performed on each face surface, to minimize recall bias in landmark selection.

5.1.1 Observer reliability results

The intra-observer and inter-observer reliability in selecting 3D landmarks from face scans, was evaluated using intraclass correlation coefficients (ICCs) described in section 3.2. The intra-observer reliability was determined for observer 1 and observer 2, independently. Thereafter, the two observer landmark selections were compared for inter-observer reliability. The ICC

calculations were implemented in the Statistical Package for the Social Sciences (SPSS) (Argyrous, 2000).

The ICC results and their corresponding 95% confidence intervals (CI) are reported for intra- and inter-observer reliability assessment. The ICC between landmark recordings obtained by observer 1 was 0.998 (95% CI: 0.997 to 0.998). The ICC between landmark recordings obtained by observer 2 was 0.999 (95% CI: 0.998-0.999). For inter-observer reliability, the ICC between landmark recordings obtained by two observers was 0.961 (95% CI: 0.952-0.968). The ICC values for the two independent observers were greater than the inter-observer ICC value. Both the intra- and the inter-observer agreements of selected landmark measurements were excellent with ICC value greater than 0.9 (Victor *et al.*, 2009).

5.1.2 Observer precision results

The mean absolute difference (MAD), described in section 3.2, was used to evaluate the intra-observer and inter-observer precision for 3D landmark identifications. The precision results for 3D facial landmark identifications are shown in Figure 5.1 and Figure 5.2.

Figure 5.1 illustrates the results of intra-observer errors of the selected landmarks based on MAD calculations. The average intra-observer error in selecting landmarks by observer 1 was 0.29 mm with range of 0.15 mm (right alare) to 0.55 mm (pogonion). For observer 1, the pogonion and the left exocanthion show slightly higher landmarking error compared to the rest of the landmarks. The average intra-observer error in selecting landmarks by observer 2 was 0.23 mm with range of 0.13 mm (pronasale) to 0.48 mm (left alare). For observer 2, the left alare and right alare indicate a slightly higher landmarking error in comparison to the rest of the landmarks. Considering observer 1 and observer 2, all the landmarks were selected with a landmarking error less than 1 mm as defined by precision levels in Table 3.1

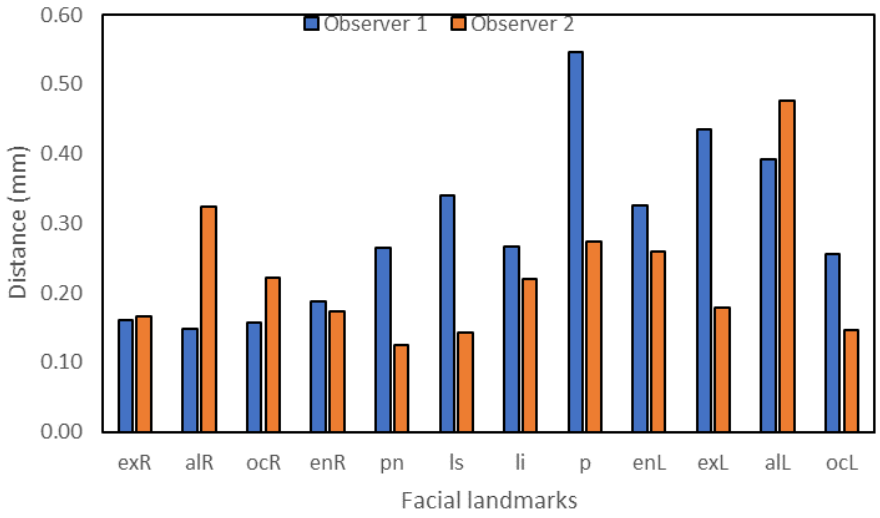


Figure 5.1: Intra-observer precision of the landmarks. The histograms indicate the MAD associated with each facial landmark.

Figure 5.2 shows the inter-observer precision of the selected facial landmarks. The average inter-observer error was 0.89 mm with a range of 0.24 mm to 1.75 mm. Seven of the 12 facial landmarks were selected with landmarking errors of less 1 mm, three facial landmarks were selected with errors between 1 mm and 1.5 mm, and two landmarks (pogonion and left alare) were selected with landmarking errors between 1.6 mm - 2 mm. No landmark had an error above 2 mm. The precision levels were defined according to Table 3.1.

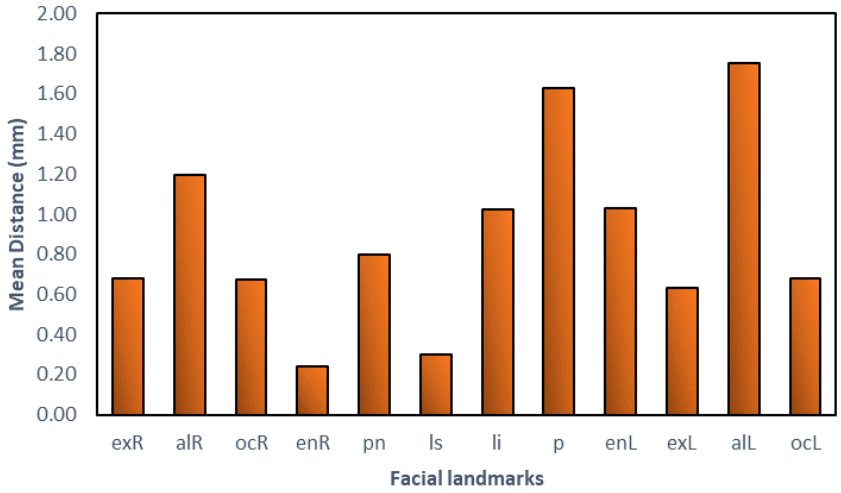


Figure 5.2: Inter-observer precision of the landmarks. The histograms show the MAD associated with each facial landmark.

5.2 Preprocessing of 3D face scans

Once the 3D face scans had been obtained from the BU-3DFE database referred to in section 4.2, the scans were preprocessed. The preprocessing stage involved trimming of the 3D face scans to remove the unwanted regions such as the hair and neck regions. Figure 5.3 illustrates an original face scan and the trimmed face scan. The face surface trimming was implemented automatically using the Scalismo software tool discussed in section 4.6.

Figure 5.3: Preprocessing: (a) the original face scan from the BU-3DFE dataset and (b) the trimmed face scan.

5.3 Rigid registration of 3D face scans

The preprocessed 3D scans needed to be placed in the common coordinate system. The set of anatomical facial landmarks discussed in section 4.5 were used as a guide to establish sparse correspondences across the scans in the training set. The training face scans were mapped, using a rigid transformation, to a mean Basel face model (BFM) (Paysan *et al.*, 2009), which acted as a reference surface. The BFM was used because it is publicly available for research and provides accurate triangulations of the face surface collected from high resolution face scans. Rigid registration of the face scans was implemented using the Scalismo software tool discussed in section 4.6. The result of this registration was a set of rigidly aligned 3D face scans.

5.4 Non-rigid registration of 3D face scans

Given a set of aligned face scans in which each face scan has its own number of vertices and triangles, the next step was to obtain dense correspondences for the 3D face scans in the training set. The aim of finding dense correspondences was to re-parameterize face scans so that they have the same number of vertices and triangulations across the training set. Dense correspondence was performed by fitting a reference face surface to each target face scan in the training set. The choice of the reference face used was again the mean face from the BFM (Paysan *et al.*, 2009). The reference face surface was fitted to each target face scan by applying a Gaussian process fitting (Gerig *et al.*, 2014) using the Scalismo software library discussed in section 4.6 in this way, dense surface deformations were obtained, which best match a target face scan to a reference face with no spatial variability. The results for the application of the non-rigid registration approach were registered face scans each with 53,149 vertices and 105,694 triangles. The registered face scans were later statistically analyzed to construct 3DMMs of faces (shape and texture) as described in section 5.5.

5.5 Statistical modelling

After establishing dense correspondences for all the face scans, the next step was to remove any misalignments of the face scans to obtain facial shape variations only. Any spatial variations due to misalignment of face scans would be modelled along with shape and texture variability when building the statistical distribution models, resulting in inaccurate models. These non-shape related spatial variations were removed by performing a similarity alignment using generalized Procrustes analysis (GPA) (Gower, 1975) implemented in Scalismo (see section 4.6).

After alignment of the 3D face scans, 3DMMs were developed following the Gaussian processes discussed in section 3.5. The 3DMMs with shape and texture information were constructed from the registered 3D face scans expressed as linear combinations of shape and texture vectors in face subspace. The full 3DMM construction pipeline is illustrated in Figure 5.4. The major parts in the 3DMM construction process are alignment of face scans, non-rigid registration of face scans based on the BFM as the reference and finally building the morphable models with shape and texture

properties from the registered face scans. In Figure 5.4, the shaded rectangles indicate the outputs of major steps of the 3DMM construction process.

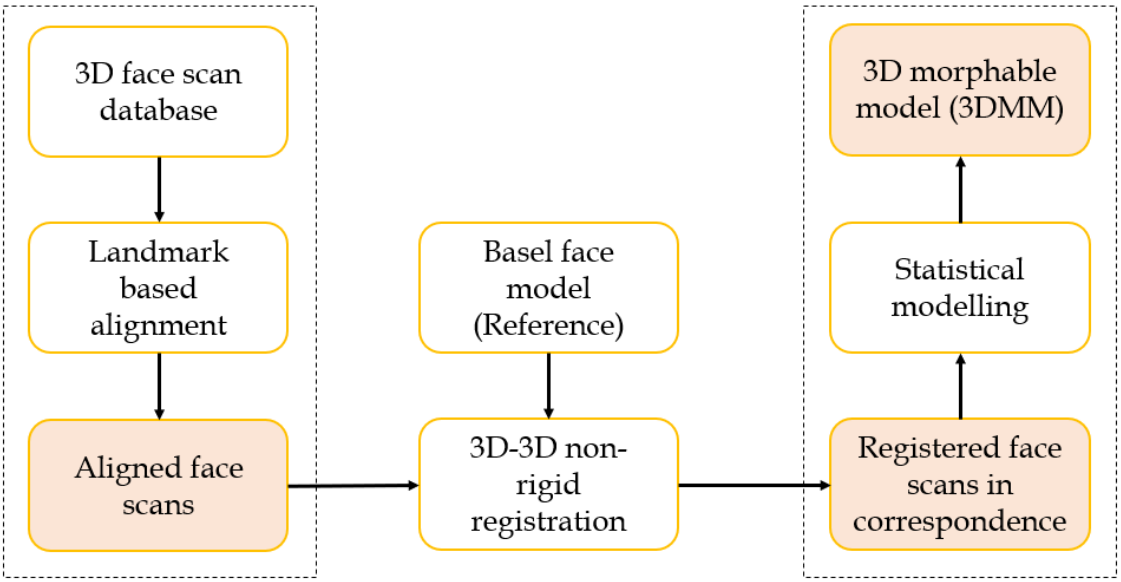


Figure 5.4: The full construction pipelineof 3DMM of human faces. The shaded rectangles show the output of each process.

5.6 3DMM Visualization

To understand the high-dimensional space of the human face shape, the new shape instances are generated from the 3DMM of the face using equation (5.1) (Gerig *et al.*, 2018).

$$X^* = \bar{X} + \sum_{i=1}^{n} \vartheta_i \sqrt{\lambda_i}\vec{v}_i, \qquad \vartheta_i \sim N(0,1) \tag{5.1}$$

where $\bar{X}$ is the mean shape of the face, ϑ_i is a shape vector with a set of parameters of a 3DMM drawn from a normal distribution of shapes, λ_i represents the eigenvalues from the basis matrix, $\vec{v}_i$ represents shape eigenvectors (principal components) from the columns of the basis matrix, and finally X^* is a new shape instance generated from linear combinations of random face shapes and the mean face shape.

The mean shape of the face and first four shape principal components from the constructed face shape model are presented in Figure 5.5. The depictions are samples from the face shape model at $\pm 3\sigma$, (where σ is the standard deviation) from the mean shape in the direction of each shape principal component. The first shape principal component ($PCs\ 1$) represents the highest

geometric variation in the dataset. The second principal component ($PCs\ 2$) represents the second highest geometric variation in the dataset and so on until the last shape principal component in the dataset.

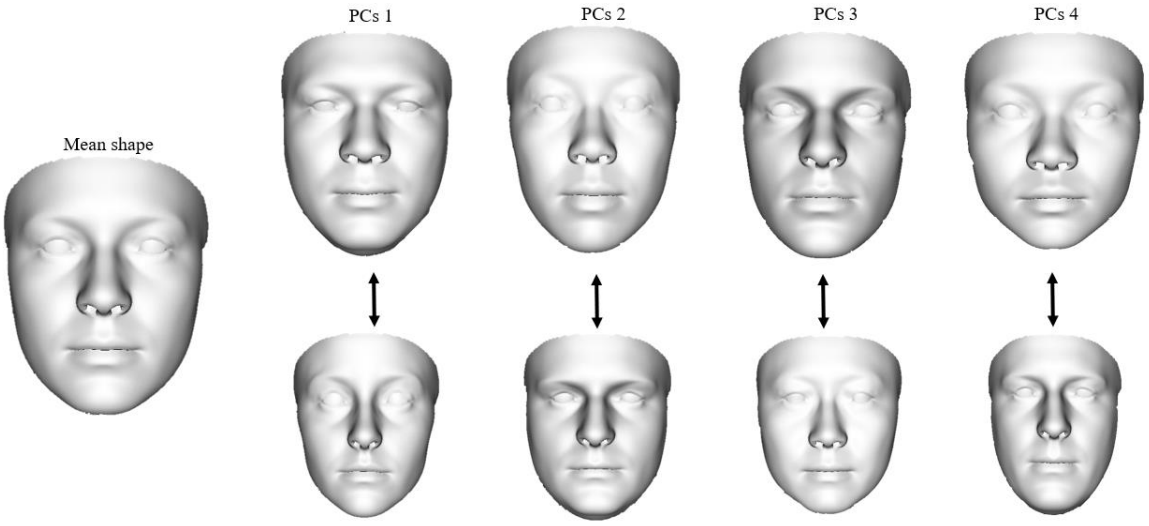


Figure 5.5: Visualizations from the shape model showing shape variation patterns.The mean shape and the first four shape principal components are displayed. The first row represents $+3$ standard deviations and the second row represents -3 standard deviations from the shape mean in the direction of each shape principal component.

The constructed 3DMM of the face contains shape and texture information. The texture features of the 3DMM are examined below. The face textures from nearby vertices are generated by interpolating color values using equation (5.2) (Gerig *et al.*, 2018).

$$C^* = \bar{C} + \sum_{i=1}^{n} \beta_i \sqrt{\lambda_i}\vec{v}_i, \qquad \beta_i \sim N(0,1) \tag{5.2}$$

where $\bar{C}$ is the mean texture of the face, β_i is a set of parameters of a 3DMM drawn from a normal distribution of textures, λ_i represents the eigenvalues from the basis matrix, $\vec{v}_i$ represents texture eigenvectors (principal components) from the columns of the basis matrix, and C^* is a new texture instance within the face population generated from a linear combination of random face surfaces and the mean texture of the face.

The mean texture of the face and the first four texture principal components of the texture model are illustrated in Figure 5.6. The first texture principal component ($PC_t\ 1$) represents the highest texture variation in the dataset. The second texture component ($PC_t\ 1$) represents the second highest texture variations in the dataset and so on. The depictions in Figure 5.6 are samples from

the texture model at $\pm 3\sigma$, (where σ is the standard deviation) from the mean texture in the direction of each texture principal component. It can be observed from Figure 5.6 that the varying face textures present in the training samples were captured by the morphable texture model. However, the face textures are smooth suggesting that the finer details of the human face like wrinkles and moles are not represented in the texture model.

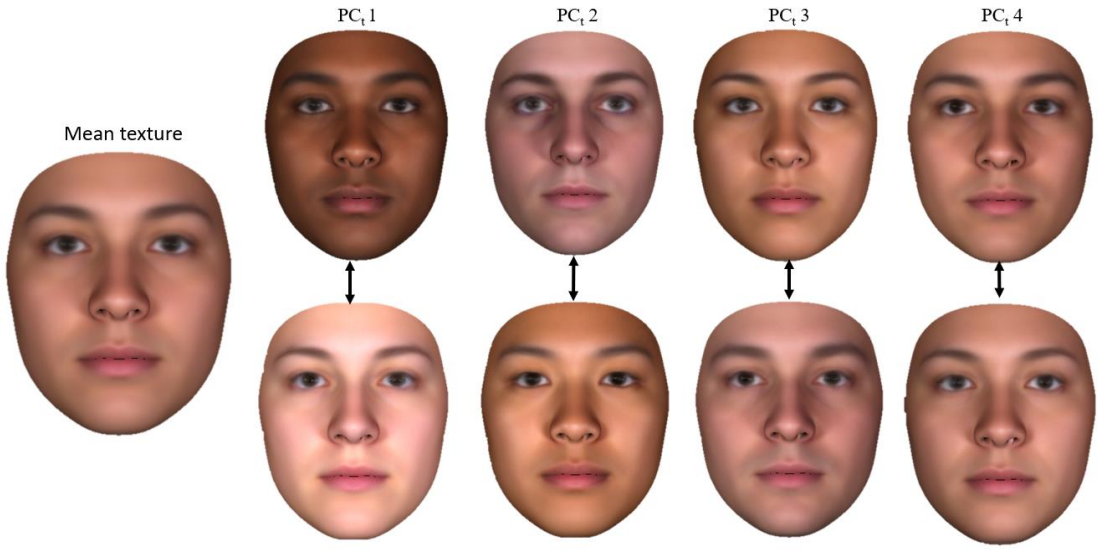


Figure 5.6: Visualization of variations from the texture model. The mean texture and the first four texture principal components are presented. The first row represents $+3$ standard deviations and the second row represents -3 standard deviations from the mean texture in the direction of each texture principal component.

5.7 Evaluation of 3DMM quality

After 3DMM construction from the 3D face scans, the quality of the 3DMM was assessed based on three evaluation metrics (generalization, specificity, and compactness) proposed by Styner *et al.* (2003). The quality assessment was based on the shape component of the 3DMM only, since the reconstruction is more focused on the geometric aspects of the human face. Sections 5.7.1, 5.7.2, and 5.7.3 detail the quantitative evaluation metrics.

5.7.1 Shape model generalization

The leave-one-out approach was used to quantify the generalization ability of the face shape model. For each iteration, a shape model was constructed from a set of training face surfaces leaving out one face shape instance. As all the training data were in correspondence, the left-out

face instance was projected into the shape model space and an approximation of the face generated. To evaluate the geometric accuracy of the estimated face, the distance between the face instance estimate and the original face instance was calculated. The accuracy metric used was the average vertex-to-vertex rms distance between the left-out face instance and the estimated face instance. The procedure was repeated until all the face surfaces in the training set were used.

Model generalization ability results are presented in Figure *5.7*, which demonstrates the generalization error plotted against shape principal components. After the first 25 principal components, the generalization of the shape model measured as rms distance was less than 1 mm. After the first 45 principal components, the distance line levels off to approximately 0.5 mm.

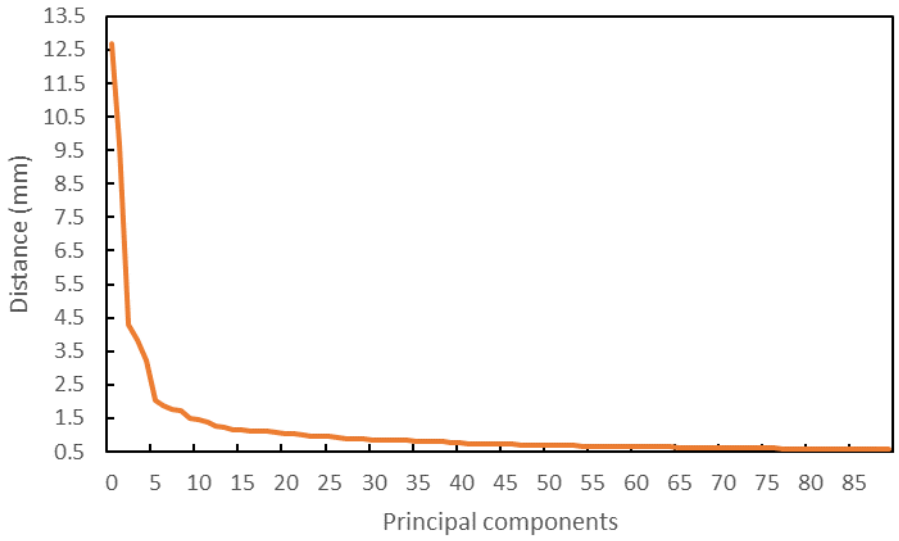


Figure 5.7: *RMS distances plotted against the numbers of principal components.*

5.7.2 Shape model specificity

The specificity of the shape model is the ability of the model to randomly generate synthetic shape instances that are similar to real shape instances present in the training set (Styner *et al.*, 2003). To evaluate model specificity, a set of 100 shape instances were randomly generated from the shape distribution of the 3DMM. The rms distance between the randomly generated shape instances and the closest face surfaces in the training set was calculated as a specificity estimate. Lower rms deviations are desirable since they indicate that the synthesized shape instances are close to the real shape instances in the training set.

Figure 5.8 shows the specificity results. The graph levels off at around 14.5 mm as the number of principal components of the shape model increases. The rms distances range from 13.2 mm to 14.5 mm.

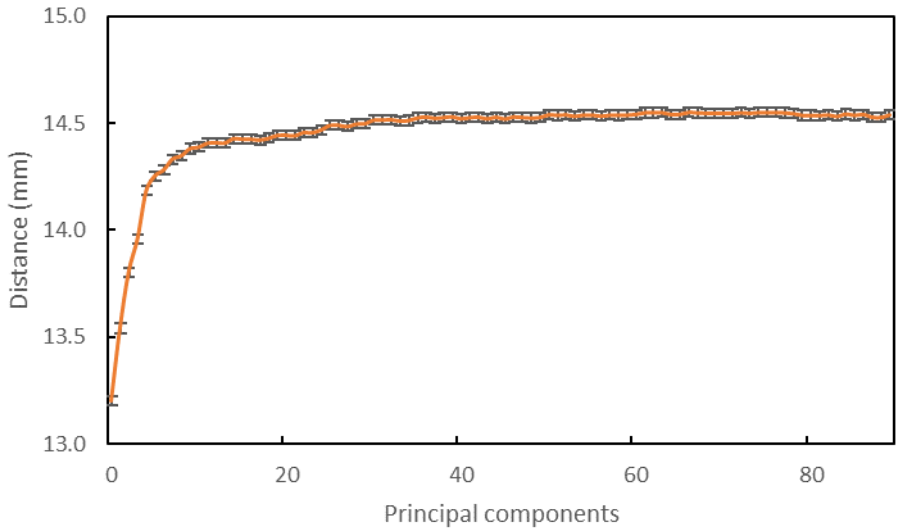


Figure 5.8: Specificity: RMS distances plotted against the number of principal components.

5.7.3 Shape model compactness

Shape model compactness indicates the percentage of variability retained by different numbers of principal components. Fewer principal components indicate that variability in shape information is captured more efficiently. To validate the compactness, the cumulative variance retained by the shape model was plotted against the number of principal components of the shape model as seen in Figure 5.9. The graph flattens as the number of shape principal components increases. Using only the first 20 shape principal components, the shape model retains more than 90% of shape variation in the training dataset. This implies that the shape model is compact since it describes the training set using a small number of principal components.

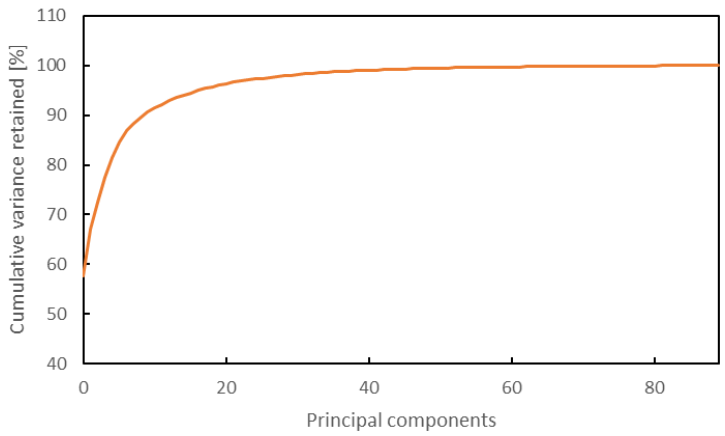


Figure 5.9: Cumulative variance plotted against number of principal components.

5.8 Discussion

The intra- and inter-observer ICC results were above 0.90 and therefore showed that landmark selection was highly reliable. The intra-observer and inter-observer errors for all the landmarks showed sub-millimeter precision and are comparable to the literature results. Mutsvangwa et al. (2011) reported a mean intra-observer error of 0.71 mm and inter-observer error of 0.86 mm using 26 facial landmarks. Aldridge et al. (2005) reported an average intra-observer error of 0.83 mm using 20 facial landmarks.

The 3D morphable model of the face was built using 3D face scans from which point-to-point correspondences across the 3D scans were established. Visual inspection of a sample of facial shapes from the 3DMM (see Figure 5.5) indicated that the model was representative of the training data. Similarly, a sample of facial textures from a 3DMM (see Figure 5.6) showed a range of skin tones which are representative of training data used.

Model generalization errors (see Figure *5.7*) were low indicating that the shape model performed well when describing data outside the training set. The specificity results (see Figure 5.8) from this study were in the range of 13.2 mm to 14.5 mm. No literature was found to which our specificity results could be compared. As presented in Figure 5.9 showing model compactness, more than 90% of the variability of the shape model is retained with only 20 principal components, suggesting that the shape model is compact since it describes the training set using a low number of principal components.

Chapter 6. 3D face reconstruction from a single 2D image using a 3DMM

The task of estimating a 3D face structure from a single facial image is challenging yet has many applications in face analysis, face tracking, face animation, and face recognition. To perform 3D face reconstruction, 3DMMs are used to provide prior knowledge for inferring a full 3D face representation from a single image. The first step involves establishing a correspondence between a 3DMM and a 2D facial image followed by computing the parameters of the image synthesized from 3DMM, which can explain the observed 2D image. Some of the parameters of the face that are optimized during 3DMM fitting include the shape (identity geometry), skin texture (reflectance), pose (camera), and illumination (lighting) parameters. These parameters may be estimated independently or in groups, including model parameters which consist of shape and texture; and imaging consisting of pose and illumination. The model and imaging parameter estimation is detailed in section 6.2.2. The analysis-by-synthesis approach is exploited to match a 3DMM to a single 2D image of the face (Blanz *et al.*, 2003). For effective 3DMM fitting, facial landmarks of the 2D image are defined.

This chapter presents approaches for estimating 3D reconstructions of human faces from single facial images using the 3D morphable model described in Chapter 5. Additionally, the observer precision of the selected 2D landmarks is described. Section 6.1 describes how the reliability and precision of the image landmarks was determined. Section 6.2 explains the steps involved in the 3DMM fitting pipeline to estimate 3D face surfaces. Section 6.3 evaluates the geometric quality of the 3D face reconstructions obtained from single 2D facial images.

6.1 Evaluation of 2D landmarking quality

Image landmarks are used as key points in 3DMM fitting methods to initialize and guide the registration procedures (Blanz *et al.*, 1999; Schönborn *et al.*, 2016). Landmarking a 2D image of the face was typically a manual task that may introduce errors. The extent of such errors was evaluated in this part of the study. Similarly to the 3D case, this was achieved by quantifying the reliability and precision in 2D landmark measurements obtained by different observers, using the intraclass correlation coefficient (ICC) and mean absolute distance (MAD), described in section 3.2.

The 12 facial landmarks described in section 4.5 were assessed. The landmarking methodology is as described in section 5.1, using 2D rather than 3D images. Two observers selected a set of landmarks by clicking on each facial image twice. A period of 24 hours was allowed to pass between the landmarking process for the two trials per observer performed on each facial image, to minimize recall bias in landmark selection. Observer reliability and precision in selecting 2D landmarks were evaluated. The landmarking errors in section 6.1.2 were classified using Table 6.1 below.

Table 6.1: 2D landmarking error ranges

Error ranges	**Precision levels**
Less than 1 pixels (px)	High precision
Between 1 px - 1.5 px	Precise
Between 1.6 px - 2 px	Moderate precision
Greater than 2 px	Imprecise

6.1.1 Observer reliability results

The ICC results and their corresponding 95% confidence intervals (CI) are reported for intra- and inter-observer reliability assessment. The ICC between landmark selections obtained by observer 1 was 0.99 (95% CI: 0.99 to 1). The ICC between landmark selections obtained by for observer 2 was 0.99 (95% CI: 0.99-1). The inter-observer ICC for different observer selections was 0.98 (95% CI: 0.98 to 1). All the ICC values are high suggesting that the results are reliable.

6.1.2 Observer precision results

The results for intra-observer precision are presented in Figure 6.1. The average intra-observer error for observer 1 was 0.9 pixels (px) in a range of 0.55 px to 1.37 px. Eight of the 12 landmarks were selected with landmarking errors of less than 1 px. The remaining four landmarks were selected with errors between 1 px and 1.5 px. The average intra-observer error in selecting 2D landmarks for observer 2 was 0.9 px in a range of 0.68 px to 1.34 px. Nine out of the 12 landmarks were selected with landmarking errors below 1 px. The remaining three landmarks were selected with errors between 1 px and 1.5 px. In both observer selections, no landmarks were selected with landmarking error of above 2 px.

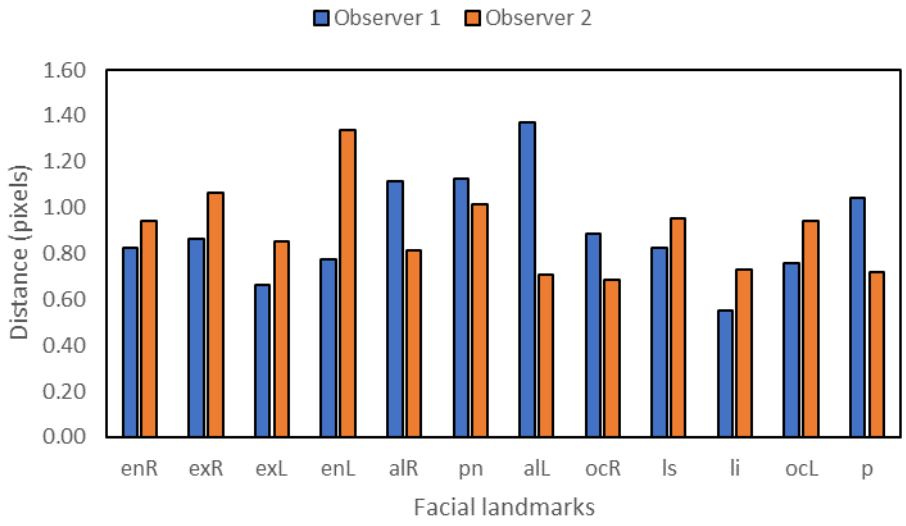


Figure 6.1*: Intra-observer precision errors of 2D facial landmarks.*

The inter-observer precision in selecting 2D landmarks is presented in Figure 6.2. The average inter-observer error was 1.25 px in the range 0.61 px to 2.76 px. Seven out of the 12 landmarks were selected with precision errors below 1 px. One landmark was selected with a 2D error between 1 px and 1.5 px. Two landmarks were selected with precision error between 1.6 px and 2 px. The remaining two landmarks (right alare and left alare) were selected with landmarking errors above 2 px.

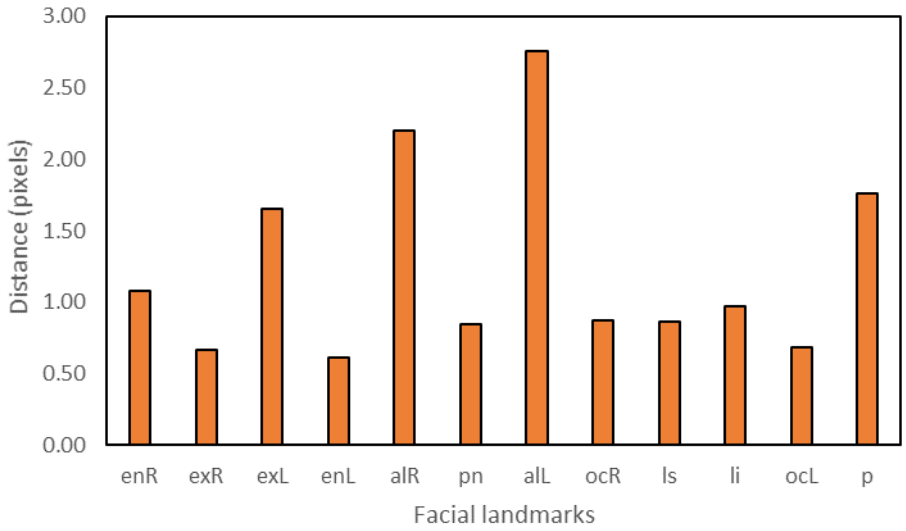


Figure 6.2: Inter-observer precision errors of 2D facial landmarks.

6.2 3DMM fitting pipeline

The 3DMM fitting approach detailed in Schönborn *et al.* (2016) was adopted for this project. The fitting method recovers a full posterior model of the face by rendering the intrinsic and extrinsic parameters of the face. The intrinsic parameters contain shape and texture information, and extrinsic parameters consist of illumination and pose parameters. The fitting approach is expressed as the optimization problem where the difference between the synthesized image and the input image is decreased. The Scalismo development toolkit (discussed in section 4.6) developed by the University of Basel was used to implement 3DMM fitting. The reconstruction framework for estimating a 3D face from a 2D image was implemented in different steps illustrated in Figure 6.3.

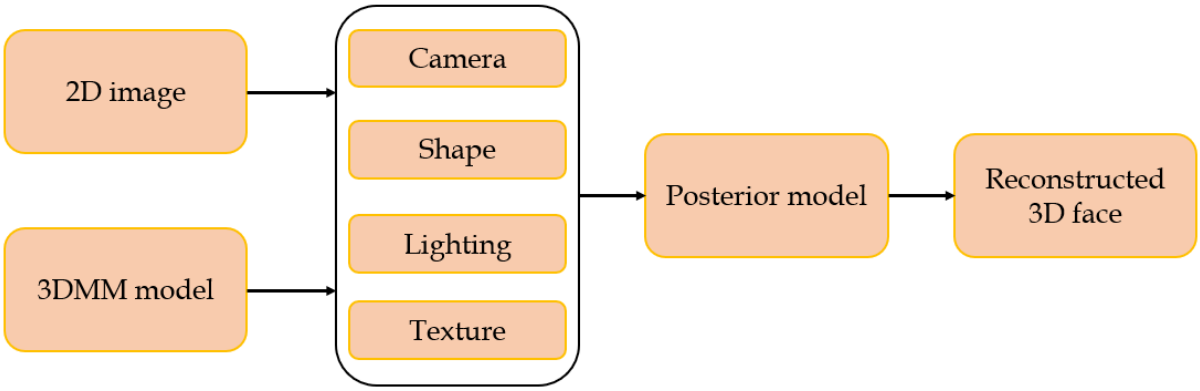


Figure 6.3: Shows 3DMM fitting steps.

The 2D data described in section 4.4 were used as input targets for model-based 3D-from-2D reconstruction process. Frontal view 2D images of the face with neutral expressions were used to perform 3D reconstruction. The 2D data provide both geometric and texture information. The geometric information guides the process of reconstructing a 3D face shape from a 2D facial image. The texture information is extracted from a 2D image and mapped onto the face geometry to provide skin texture. On each 2D facial image, 12 facial landmarks were manually selected using the interactive tool. The selected landmarks correspond to the set of facial landmarks identified during the 3DMM construction phase presented in Chapter 5. The annotated landmarks were used for initialization of the fitting algorithm. The observer reliability and precision in selecting 2D landmarks was detailed in section 6.1.

6.2.1 The Fitting method

The Markov Chain Monte Carlo (MCMC) approach proposed by Schönborn *et al.* (2016) was used to fit a 3DMM to 2D image data. The method approximates the posterior distribution model given an observed image and a 3DMM prior. For this work, 12 image landmarks which are in the same position as the landmarks used in the training data were used for initialization and guiding the 3DMM fitting strategy. The goal of this approach is to decrease the difference between the observed image and the synthetic image. Using a single 2D image and a 3DMM, the fitting algorithm optimizes the geometric parameters (pose and shape) and scene parameters (illumination and texture) to produce a 3D face representation. In section 6.2.2 below, the pose, shape, texture, and illumination parameters are explained in detail.

6.2.2 Parameter estimation

Pose parameter estimation: The estimation of the camera parameters that align the 3DMM to the 2D image were guided by a set of landmarks. The alignment of the 3DMM and the 2D image was accomplished by applying a rigid transformation T_r and a perspective projection T_p. The rigid transformation corrects for rotation and translation effects, while the perspective projection maps the 3D coordinates X_{3D} of the model to 2D coordinates $X_{2D} = (x, y)^T$ in the image plane. The projections are defined in equation (6.1) below (Hu *et al.*, 2017).

$$T_r = RX_{3D} + t$$
$$T_p \rightarrow X_{2D}, \qquad x = O_x + f\frac{w_x}{w_z}, \qquad y = O_y - f\frac{w_y}{w_z} \tag{6.1}$$

where R is a rotation matrix, t denotes a translation, f represents focal length, (O_x, O_y) denotes the center of the 2D image plane, and (w_x, w_y, w_z) represent camera-centered coordinates of X_{3D}.

The camera parameters ρ were solved by minimizing the distance between the image landmarks X_{2D}^I of the input image and the reconstructed landmarks X_{2D}^M from the 3DMM as illustrated in equation (6.2) (Hu *et al.*, 2017).

$$min \sum ||X_{2D}^I - X_{2D}^M(\rho)||^2 \tag{6.2}$$

Shape parameters estimation: The shape parameters α were estimated using the shape model prior s by aligning the 3DMM to the input 2D image. The MCMC algorithm was applied to fit a shape prior to a single 2D image. The algorithm optimizes α using a cost function that minimizes the distance between the reconstructed model landmarks X_{3D}^{M} and the observed 3D landmarks X_{3D}^{I} as defined in equation (6.3) (Hu *et al.*, 2017). The 2D coordinates X_{2D}^{I} of the input image were back projected into 3D space to give 3D coordinates X_{3D}^{I}.

$$min \sum ||X_{3D}^{I} - X_{3D}^{M}(\alpha)||^2 \tag{6.3}$$

Texture parameters estimation: The texture parameters were approximated using the texture model constructed from textures of the faces. Similar to shape parameter estimation, the texture parameters were optimized based on the cost function.

6.2.3 Posterior face model

The result of fitting a 3DMM to the target image was a posterior model of faces. The posterior face model is a distribution of face variations for a given class of faces. The posterior model was computed based on the Metropolis-Hastings (MH) algorithm and is defined as the product of the likelihood function and the statistical prior model. Due to its nature, the posterior face model contains many face samples.

6.2.4 Reconstructed 3D face

The posterior distribution model was sampled to obtain a 3D face shape reconstruction. The MH method was applied to generate the best sample, which was the mean of the posterior distribution model. Figure 6.4 presents the facial reconstructions from input 2D images. The obtained face reconstructions from single images look realistic from a visual inspection perspective. The geometric surfaces of the 3D face reconstructions were evaluated by comparing them with the associated ground-truth 3D face scans.

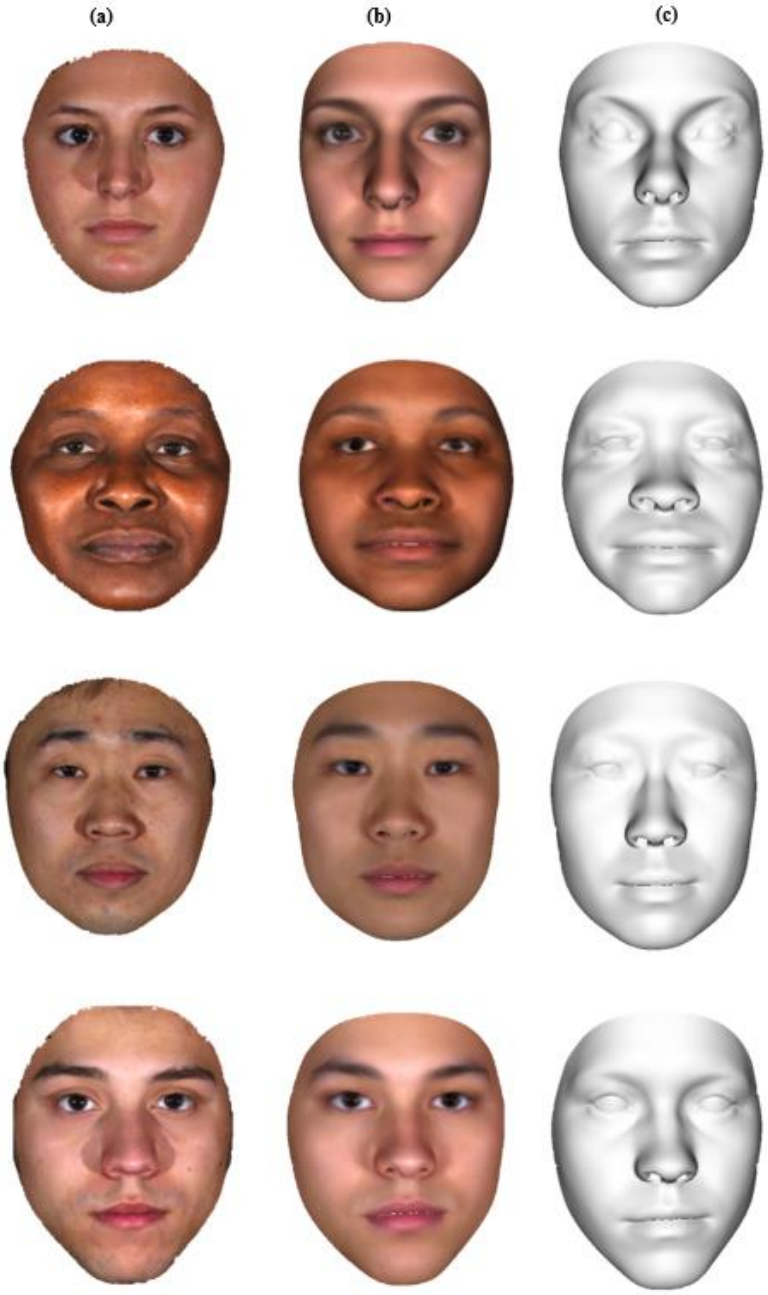


Figure 6.4: Facial reconstructions from input 2D images.The first column from left indicates single input images(a), the second column (b) shows textured face reconstructions and the third column (c) represents geometric face reconstructions.

6.3 Evaluating the geometric accuracy of 3D face reconstructions

The geometric quality of the 3D face reconstructions obtained from single 2D facial images was evaluated by comparison with ground-truth 3D face scans.

The 3D datasets working as ground-truth face scans to evaluate the geometric accuracy of the 3D face estimates were explained in section 4.3. When building and validating a 3DMM, data can be split in many ways following an 80%-20% rule or 90%-10% rule (Dua *et al.*, 2009). For this study, the data were split based on a 90%-10% rule, resulting in10 face scans in the validation set since the total data size was 100 face scans. Each reconstructed face surface was generated using a 3DMM working as prior and a single 2D image of the face working as in input (see section 6.2). A rigid alignment was performed to register the face surface estimate to the corresponding ground-truth face scan. Each face surface estimate was compared to the corresponding ground-truth face scan in a surface-to-surface comparison procedure. The evaluation metrics considered for this assessment were the Hausdorff distance and root mean square error described in sections 3.8.2 and 3.8.4, respectively.

6.3.1 Hausdorff distance

After rigid alignment of the face surfaces, the surface-to-surface distances between the two aligned faces were computed by applying Hausdorff distance and modified Hausdorff distance (mHD) functions. The details of Hausdorff and modified Hausdorff distances are presented in section 3.8.2.

Table 6.2 shows the results of geometric surface comparisons between the face shape reconstructions and the ground-truth face shapes. The average modified Hausdorff distance across the 10 pairs of aligned face surfaces (reconstructed face & ground-truth face) was 2.64 mm, with a range of 1.93 mm to 4.98 mm. The mean Hausdorff distance across the 10 pairs of aligned face surfaces was 13.86 mm ranging from 10.35 mm to 17.0 mm.

Table 6.2: Surface-to-surface distances between the reconstructed faces and ground-truth faces

Face ID	Hausdorff distance (mm)	modified Hausdorff distance (mm)
mesh-F0051	13.13	2.26
mesh-F0052	14.23	2.28
mesh-F0054	14.69	2.07
mesh-F0055	17.00	2.71
mesh-M0040	10.93	2.35
mesh-M0042	10.35	2.23
mesh-M0043	12.52	1.93
mesh-M0044	14.32	3.38
mesh-F0056	14.61	2.24
mesh-M0039	16.87	4.98

6.3.2 Root mean square errors

The global root mean square (RMS) errors between the face shape reconstruction and the corresponding ground-truth face scan were computed using the Amira software tool version 6.2.0 (Visage Imaging: http://www.vsg3d.com) described in section 4.6. Additionally, the heat maps between the face surface reconstruction and the ground-truth face scan were generated using Amira software, allowing for a visualisation of the geometric reconstruction errors. The heatmaps show the surface magnitudes for similarity and dissimilarity between the registered face surfaces using colour gradients.

The results of face shape comparisons between the surface reconstructions and the ground-truth face scans are illustrated in Table 6.3. The average RMS error was 2.99 mm ranging from 2.25 mm to 5.45 mm.

Table 6.3: Surface to surface comparisons across registered face surfaces

Face ID	RMS deviation (mm)
mesh-F0051	2.37
mesh-F0052	2.38
mesh-F0054	3.27
mesh-F0055	2.59
mesh-M0040	3.17
mesh-M0042	3.40
mesh-M0043	2.25
mesh-M0044	2.70
mesh-F0056	2.27
mesh-M0039	5.45

The color-coded maps between the 3D face surface reconstructions and the ground-truth 3D face scans are illustrated in Figure 6.5 and Figure 6.6. Considering regions of the face with features that are mainly affected by FAS, namely the eyes, the mid face, the upper lip, and the philtrum, the upper lip, and the philtrum show high facial surface variation as represented by changes in color. The eye regions and mid face areas show less variation.

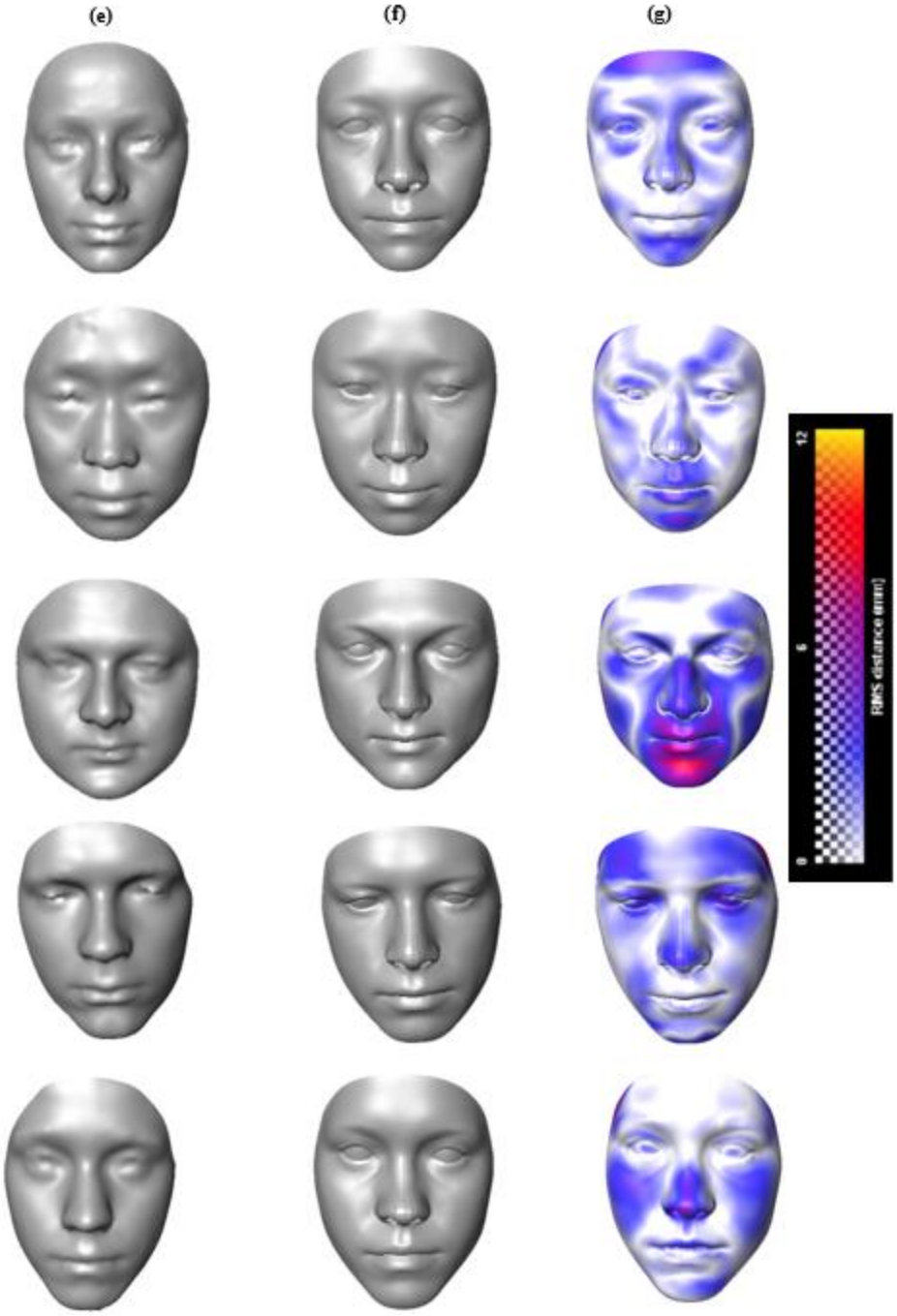


Figure 6.5: Visualization of geometric comparisons of the first 5 face surfaces. The first column (e) indicates ground-truth face scans. The second column (f) represents the reconstructed face shape. The third column (g) shows heatmaps (color-coded surface) between the ground-truth face and its associated face reconstruction.

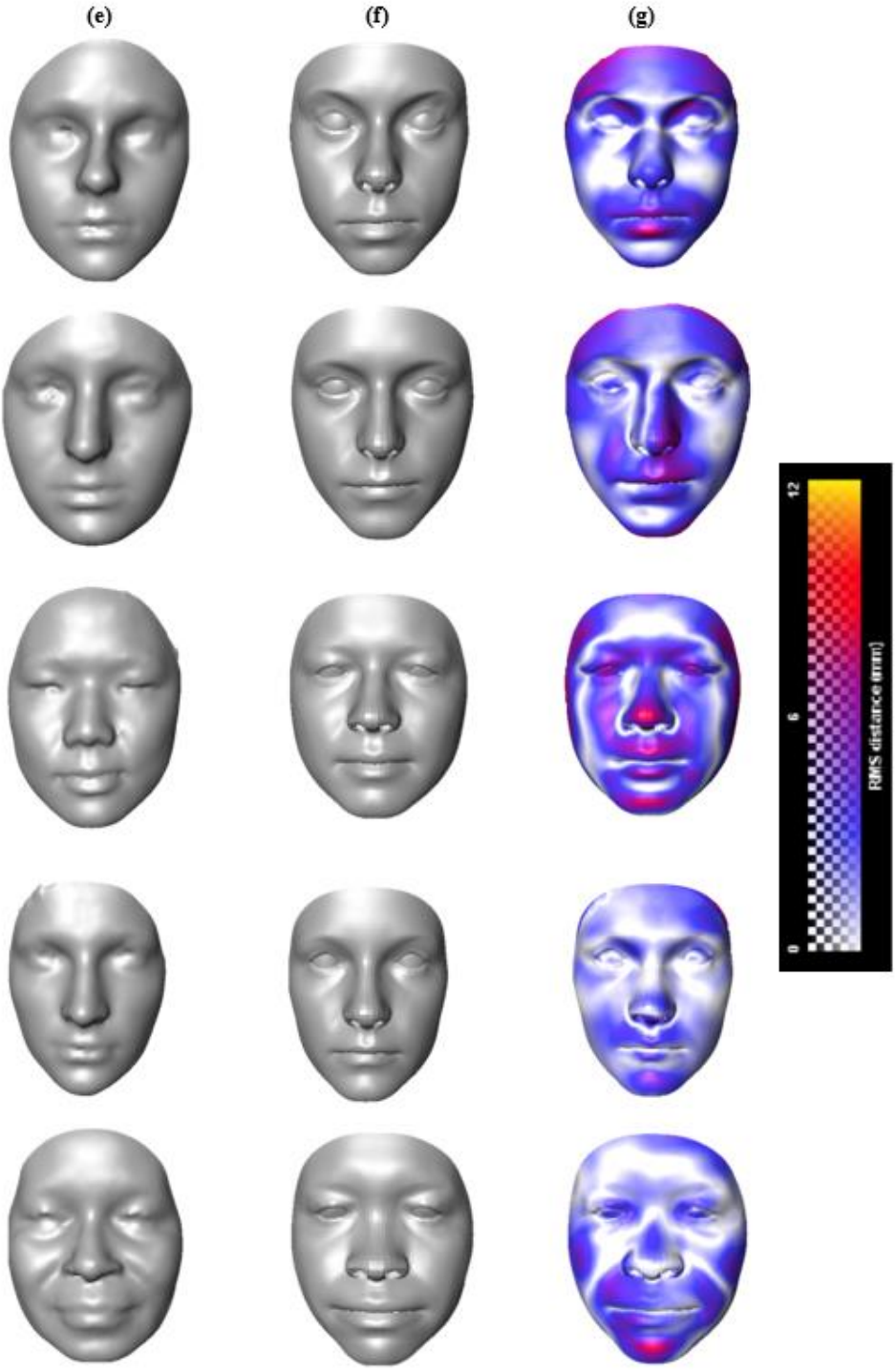


Figure 6.6: Visualization of geometric comparisons of the second 5 face surfaces. The first column (e) indicates ground-truth face scans. The second column (f) represents the reconstructed face shape. The third column (g) shows heatmaps (color-coded surface) between the ground-truth face and its associated face reconstruction.

6.4 Discussion

The chapter detailed the development of a framework for estimating a 3D face reconstruction from a single 2D facial image using a prior 3DMM. The quality of 2D landmark selections was assessed by determining observer reliability and precision. The ICC results suggest that the 2D landmark selections were reliable. The intra-observer and inter-observer errors of the selected 2D landmarks were expressed in pixels and therefore cannot be compared to 3D landmarks described in section 5.1 since they have different scaling units. Furthermore, a framework for estimating the 3D facial surface from a single 2D facial image was developed. The framework was built in an open source software, Scalismo and then used to estimate 3D face reconstruction using a 3DMM to fit a single 2D image of the face. It can be visually observed from Figure 6.4 that the fitting method performed well at approximating 3D facial surface reconstructions from single 2D images of the faces.

The geometric accuracy between the face shape reconstructions and the ground-truth face scans was evaluated both qualitatively and quantitatively. From visual observations of the reconstruction errors in Figure 6.5 and Figure 6.6, the mid face and the regions around the eyes show lower levels of surface variation as represented by the heatmaps, whereas the upper lip and the philtrum areas show greater surface variation. The lower levels of surface variation around the eyes and the midface could be attributed to ease of identifying landmarks in those regions.

The average surface-to-surface distance as measured using rms between the reconstructed face shapes and the corresponding ground-truth face scans was 2.99 mm (Table 6.3). These results are comparable to those found in the literature. Zollhofer *et al.* (2011) compared a reconstructed face surface obtained from the Kinect sensor image to a ground-truth face scan reporting an average deviation of about 2 mm. Feng *et al.* (2018) reported a root mean square error of 2.83 mm from surface comparisons between reconstructed 3D faces and the corresponding ground-truth 3D face scans.

The surface reconstruction errors may be attributed to a small number of landmarks used to guide the 3DMM fitting. Furthermore, the surface differences could imply that during the 3DMM fitting phase of the reconstruction process, a 3DMM prior did not completely capture all geometric cues in the input 2D image of the face. Here, a geometric cue is defined as the information contained in

a 2D image of the face such as configurational (relative layout of the features) information or contours. To improve on the surface reconstruction performance, future work could consider using multi-view 2D images of the face to provide more geometric cues during 3DMM fitting.

Chapter 7. General Conclusions and Recommendations

This research study aimed at developing and evaluating a framework for estimating a 3D face from a 2D image of the face based on a 3DMM of the human faces. This was achieved by first building a 3DMM from a set of registered 3D face scans. The parameters of the face contained in a 3DMM were then fitted to a single 2D image to produce a 3D face reconstruction. Additionally, the geometric accuracy of the face shape reconstructions was evaluated against ground-truth face scans using distance metrics.

7.1 3DMM construction and validation

The 3DMM of the human face was successfully developed from registered 3D face scans and evaluated for generalization, specificity, and compactness. The lowest generalization error of the constructed shape model was 0.5 mm which suggests that the face shape model describes the unseen face instances well. The shape model evaluation results for generalization compare well with other results found in the literature. The specificity values of the shape model were in range 13.2 mm to 14.5 mm. No comparable results for specificity were found in the literature. Considering compactness evaluation results, more than 90% of the variability in the shape model of the face was explained with the first 20 shape principal components and this compares well with the results discussed in the literature. For example, Booth *et al.* (2018) reported that using the first 40 shape principal components explained more than 90% of variability in the training set. Overall, the 3DMM construction and validation phases were successful.

7.2 3DMM fitting and evaluation

A 3D face reconstruction was estimated from a 2D image based on the 3DMM. From visual inspection, the reconstructed face surfaces looked realistic. The average modified Hausdorff distance across aligned face surfaces (reconstructed face & ground truth) was 2.64 mm. The mean Hausdorff distance across registered face surfaces was 13.86 mm. Hausdorff distance metrics have been used in the literature to study other anatomical structures but no literature was found for human faces. The average RMS error between the face shape estimations and the ground-truth face scans was 2.99 mm. The RMS error was slightly higher than results reported in the literature.

7.3 Applications of the 3DMM of the human face

The 3DMM reduces the search space for fitting algorithms by providing shape and texture priors during estimation of 3D reconstructions of the face from 2D facial images. The 3DMM produces personalized models for medical applications such as diagnosis of facial syndromes, surgical planning and assessment. The preliminary results show the potential of the reconstruction framework to be used for identification of facial conditions relating to fetal alcohol syndrome.

7.4 Application of 3D facial reconstructions to FAS diagnosis

The distinctive FAS facial phenotype, includes a smooth philtrum, shortened palpebral fissures and thin upper lip vermillion (Suttie *et al*., 2013). These facial features guide clinicians identifying individuals with FAS. The project aimed at reconstructing 3D facial geometries with a particular focus on assessing if the regions representing the FAS facial phenotype are faithfully reconstructed. To assess reconstruction accuracy, the 3D facial reconstructions were compared with the corresponding ground truth facial shapes obtaining a global reconstruction accuracy of 2.99 mm based on surface to surface distance comparisons. From the shape visualizations, it was observed that regions of the face around the eyes (palpebral fissures) were estimated with high accuracy as seen by colour gradient from heat maps (see Figure 6.5 & Figure 6.6). Additionally, it was noted that facial areas such as the upper lip vermillion and philtrum which are essential for FAS facial diagnosis were moderately reconstructed (see Figure 6.5 and Figure 6.6). The reconstruction errors around the philtrum and upper lip vermillion areas could be reduced by adding more landmarks in those regions. Generally, the reconstruction method presented in this study shows potential as a strategy for evaluating facial characteristics related to FAS using 3D facial reconstructions estimated from single 2D images.

7.5 Limitations and future work

Manual landmarking was performed in this research project. This tends to be time-consuming for large datasets and is error prone. Automating the landmarking process may reduce such errors.

The 3D morphable model was constructed from 90 registered face scans. Future work should focus on reducing the reconstruction errors to acceptable clinical standards by collecting and analyzing

larger datasets. Including more training data, especially from underrepresented populations, would broaden the applicability of the developed morphable models.

Additionally, future work should consider estimating a 3D face using multi-view 2D images of the face to increase the information available to perform a 3D-from-2D reconstruction.

Reference List

Aase, J. M. (1994). Clinical recognition of FAS: difficulties of detection and diagnosis. *Alcohol Health and Research World,vol. 18*(1),pp. 5.

Aldridge, K., Boyadjiev, S. A., Capone, G. T., DeLeon, V. B., & Richtsmeier, J. T. (2005). Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *American Journal of Medical Genetics Part A,vol. 138*(3),pp. 247-253.

Allen, B., Curless, B., & Popovic, Z. (2003). The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics,vol. 22*(3),pp. 587-594.

Amberg, B., Romdhani, S., & Vetter, T. (2007). *Optimal step nonrigid ICP algorithms for surface registration.* Paper Presented at the Computer Vision and Pattern Recognition.,pp. 1-8.

Arellano, C., & Dahyot, R. (2012). *Shape model fitting algorithm without point correspondence.* Paper Presented at the 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO),pp. 934-938.

Argyrous, G. (2000). *Statistics for social and health research: with a guide to SPSS*: Sage.

Astley, S. J. (2006). Comparison of the 4-digit diagnostic code and the Hoyme diagnostic guidelines for fetal alcohol spectrum disorders. *Pediatrics,vol. 118*(4),pp. 1532-1545.

Astley, S. J., & Clarren, S. K. (1995). A fetal alcohol syndrome screening tool. *Alcoholism, Clinical and Experimental Research,vol. 19*(6),pp. 1565-1571.

Astley, S. J., & Clarren, S. K. (2000). Diagnosing the full spectrum of fetal alcohol-exposed individuals: introducing the 4-digit diagnostic code. *Alcohol and Alcoholism,vol. 35*(4),pp. 400-410.

Astley, S. J., & Clarren, S. K. (2001). Measuring the facial phenotype of individuals with prenatal alcohol exposure: correlations with brain dysfunction. *Alcohol and Alcoholism,vol. 36*(2),pp. 147-159.

Astley, S. J., Magnuson, S. I., Omnell, L. M., & Clarren, S. K. (1999). Fetal alcohol syndrome: changes in craniofacial form with age, cognition, and timing of ethanol exposure in the macaque. *Teratology,vol. 59*(3),pp. 163-172.

Aubert, B., Vazquez, C., Cresson, T., Parent, S., & De Guise, J. (2016). *Automatic spine and pelvis detection in frontal X-rays using deep neural networks for patch displacement learning.* Paper Presented at the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI),pp. 1426-1429.

Bacivarov, I., & Corcoran, P. M. (2009). *Facial expression modeling using component AAM models—Gaming applications.* Paper Presented at the Games Innovations Conference, 2009.,pp. 1-16.

Bentsianov, B., & Blitzer, A. (2004). Facial anatomy. *Clinics in Dermatology,vol. 22*(1),pp. 3-13.

Berlinet, A., & Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*: Springer Science & Business Media.

Besl, P. J., & Mckay, N. D. (1992). A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 14*(2),pp. 239-256.

Blanz, V., & Vetter, T. (1999). *A morphable model for the synthesis of 3D faces.* Paper Presented at the Siggraph,vol. *99*,pp. 187-194.

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 25*(9),pp. 1063-1074.

Booth, J., Antonakos, E., Ploumpis, S., Trigeorgis, G., Panagakis, Y., & Zafeiriou, S. (2017). *3D Face Morphable Models "In-the-Wild".* Paper Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),pp. 5464-5473.

Booth, J., Roussos, A., Ponniah, A., Dunaway, D., & Zafeiriou, S. (2018). Large Scale 3D Morphable Models. *International Journal of Computer Vision,vol. 126*(2-4),pp. 233-254.

Burgstaller, B., & Pillichshammer, F. (2009). The average distance between two points. *Bulletin of the Australian Mathematical Society,vol. 80*(3),pp. 353-359.

Celiktutan, O., Ulukaya, S., & Sankur, B. (2013). A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing,vol. 2013*(1),pp. 13-40.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development,vol. 7*(3),pp. 1247-1250.

Champleboux, G., Lavallee, S., Szeliski, R., & Brunie, L. (1992). *From accurate range imaging sensor calibration to accurate model-based 3D object localization.* Paper Presented at the Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,pp. 83-89.

Chang, J. B., Small, K. H., Choi, M., & Karp, N. S. (2015). Three-dimensional surface imaging in plastic surgery: foundation, practical applications, and beyond. *Plastic and Reconstructive Surgery,vol. 135*(5),pp. 1295-1304.

Cheng, S. Y., Marras, I., Zafeiriou, S., & Pantic, M. (2017). Statistical non-rigid ICP algorithm and its application to 3D face alignment. *Image and Vision Computing,vol. 58*,pp. 3-12.

Chudley, A. E., Kilgour, A. R., Cranston, M., & Edwards, M. (2007). *Challenges of diagnosis in fetal alcohol syndrome and fetal alcohol spectrum disorder in the adult.* Paper Presented at the American Journal of Medical Genetics Part C: Seminars in Medical Genetics,vol. *145*(3),pp. 261-272.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., & Ranzuglia, G. (2008). *Meshlab: an open-source mesh processing tool.* Paper Presented at the Eurographics Italian chapter conference,vol. *2008*,pp. 129-136.

Clarren, S. K., Sampson, P. D., Larsen, J., Donnell, D. J., Barr, H. M., Bookstein, F. L., . . . Streissguth, A. P. (1987). Facial effects of fetal alcohol exposure: assessment by photographs and morphometric analysis. *American Journal of Medical Genetics,vol. 26*(3),pp. 651-666.

Cook, C. (2000). A Review of Intraclass Correlation.

Cootes, T. F., & Taylor, C. J. (2001). *Statistical models of appearance for medical image analysis and computer vision.* Paper Presented at the Medical Imaging 2001: Image Processing,vol. *4322*,pp. 236-249.

Dai, H., Pears, N., Smith, W. A., & Duncan, C. (2017). *A 3d morphable model of craniofacial shape and texture variation.* Paper Presented at the Proceedings of the IEEE International Conference on Computer Vision,pp. 3085-3093.

Danielsson, P. E. (1980). Euclidean Distance Mapping. *Computer Graphics and image processing,vol. 14*(3),pp. 227-248.

Douglas, T. S., Martinez, F., Meintjes, E. M., Vaughan, C. L., & Viljoen, D. L. (2003). Eye feature extraction for diagnosing the facial phenotype associated with fetal alcohol syndrome. *Medical and Biological Engineering and Computing,vol. 41*(1),pp. 101-106.

Dua, S., Singh, H., & Thompson, H. W. (2009). Associative classification of mammograms using weighted rules. *Expert systems with applications,vol. 36*(5),pp. 9250-9259.

Dubuisson, M. P., & Jain, A. K. (1994). *A modified Hausdorff distance for object matching.* Paper Presented at the Proceedings of 12th International Conference on Pattern Recognition,vol. *1*,pp. 566-568.

Farkas, L. G. (1994). *Anthropometry of the Head and Face*: Raven Pr.

Feng, Z.-H., Huber, P., Kittler, J., Hancock, P., Wu, X.-J., Zhao, Q., . . . Rätsch, M. (2018). *Evaluation of dense 3D reconstruction from 2D face images in the wild.* Paper Presented at the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),pp. 780-786.

Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., & Vetter, T. (2018). *Morphable face models-an open framework.* Paper Presented at the Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on,pp. 75-82.

Gerig, T., Shahim, K., Reyes, M., Vetter, T., & Lüthi, M. (2014). *Spatially varying registration using gaussian processes.* Paper Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention,pp. 413-420.

Gower, J. C. (1975). Generalized Procrustes Analysis. *Psychometrika,vol. 40*(1),pp. 33-51.

Halazonetis, D. J. (2004). Morphometrics for cephalometric diagnosis. *American Journal of Orthodontics and Dentofacial Orthopedics,vol. 125*(5),pp. 571-581.

Hammond, P., Hutton, T. J., Allanson, J. E., Campbell, L. E., Hennekam, R. C., Holden, S., . . . Winter, R. M. (2004). 3D analysis of facial morphology. *American Journal of Medical Genetics. Part A,vol. 126A*(4),pp. 339-348.

Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika,vol. 57*(1),pp. 97-110.

Heimann, T., & Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis,vol. 13*(4),pp. 543-563.

Honrado, C. P., & Larrabee, W. F., Jr. (2004). Update in three-dimensional imaging in facial plastic surgery. *Current Opinion in Otolaryngology & Head and Neck Surgery,vol. 12*(4),pp. 327-331.

Hu, G., Yan, F., Kittler, J., Christmas, W., Chan, C. H., Feng, Z., & Huber, P. (2017). Efficient 3D morphable face model fitting. *Pattern Recognition,vol. 67*,pp. 366-379.

Huang, J., Jain, A., Fang, S., & Riley, E. P. (2005). *Using facial images to diagnose fetal alcohol syndrome (FAS).* Paper Presented at the International Conference on Information Technology: Coding and Computing (ITCC'05),vol. *2*,pp. 66-71

Huang, X., Paragios, N., & Metaxas, D. N. (2006). Shape registration in implicit spaces using information theory and free form deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 28*(8),pp. 1303-1318.

Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing Images Using the Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 15*(9),pp. 850-863.

Jones, K. L. (2014). Fetal alcohol syndrome. In (Vol. 43, pp. 82-82).

Kabus, S., Netsch, T., Fischer, B., & Modersitzki, J. (2004). *B-spline registration of 3D images with Levenberg-Marquardt optimization.* Paper Presented at the Medical Imaging 2004: Image Processing,vol. *5370*,pp. 304-314.

Largo, R. D., Wettstein, R., Fulco, I., Tremp, M., Schaefer, D. J., Gubisch, W., & Haug, M. D. (2013). Three-dimensional laser surface scanning in rhinosurgery. *Facial Plastic Surgery,vol. 29*(02),pp. 116-120.

Lekakis, G., Claes, P., S Hamilton, G., & W Hellings, P. (2016). Three-Dimensional Surface Imaging and the Continuous Evolution of Preoperative and Postoperative Assessment in Rhinoplasty. *Facial Plastic Surgery,vol. 32*,pp. 88-94.

Lüthi, M., Gerig, T., Jud, C., & Vetter, T. (2017). Gaussian process morphable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 40*(8),pp. 1860-1873.

Lüthi, M., Jud, C., & Vetter, T. (2013). *A unified approach to shape model fitting and non-rigid registration.* Paper Presented at the International workshop on Machine Learning in Medical Imaging,pp. 66-73.

Mani, V. R. S., & rivazhagan, D. S. (2013). Survey of Medical Image Registration. *Journal of Biomedical Engineering and Technology,vol. 1*(2),pp. 8-25.

May, P. A., Marais, A. S., de Vries, M. M., Kalberg, W. O., Buckley, D., Hasken, J. M., . . . Hoyme, H. E. (2016). The continuum of fetal alcohol spectrum disorders in a community in South Africa: Prevalence and characteristics in a fifth sample. *Drug and Alcohol Dependence,vol. 168*,pp. 274-286.

Meintjes, E. M., Douglas, T. S., Martinez, F., Vaughan, C. L., Adams, L. P., Stekhoven, A., & Viljoen, D. (2002). A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome. *Medical Engineering and Physics,vol. 24*(10),pp. 683-689.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics,vol. 21*(6),pp. 1087-1092.

Moore, E. S., Ward, R. E., Jamison, P. L., Morris, C. A., Bader, P. I., & Hall, B. D. (2001). The subtle facial signs of prenatal exposure to alcohol: an anthropometric approach. *Journal of Pediatrics,vol. 139*(2),pp. 215-219.

Moore, E. S., Ward, R. E., Jamison, P. L., Morris, C. A., Bader, P. I., & Hall, B. D. (2002). New perspectives on the face in fetal alcohol syndrome: What anthropometry tells us. *American Journal of Medical Genetics,vol. 109*(4),pp. 249-260.

Moore, E. S., Ward, R. E., Wetherill, L. F., Rogers, J. L., Autti-Ramo, I., Fagerlund, A., . . . Cifasd. (2007). Unique facial features distinguish fetal alcohol syndrome patients and controls in diverse ethnic populations. *Alcoholism, Clinical and Experimental Research,vol. 31*(10),pp. 1707-1713.

Mutsvangwa, T., & Douglas, T. S. (2007). Morphometric analysis of facial landmark data to characterize the facial phenotype associated with fetal alcohol syndrome. *Journal of Anatomy,vol. 210*(2),pp. 209-220.

Mutsvangwa, T. E., Meintjes, E. M., Viljoen, D. L., & Douglas, T. S. (2010). Morphometric analysis and classification of the facial phenotype associated with fetal alcohol syndrome in 5- and 12-year-old children. *American Journal of Medical Genetics. Part A,vol. 152A*(1),pp. 32-41.

Mutsvangwa, T. E., Smit, J., Hoyme, H. E., Kalberg, W., Viljoen, D. L., Meintjes, E. M., & Douglas, T. S. (2009). Design, construction, and testing of a stereo-photogrammetric tool for the diagnosis of fetal alcohol syndrome in infants. *IEEE Transactions on Medical Imaging,vol. 28*(9),pp. 1448-1458.

Mutsvangwa, T. E., Veeraragoo, M., & Douglas, T. S. (2011). Precision assessment of stereo-photogrammetrically derived facial landmarks in infants. *Ann Anat,vol. 193*(2),pp. 100-105.

Oliveira, F. P., & Tavares, J. M. (2014). Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering,vol. 17*(2),pp. 73-93.

Ort, R., Metzler, P., Kruse, A. L., Matthews, F., Zemann, W., Grätz, K. W., & Luebbers, H.-T. (2012). The reliability of a three-dimensional photo system-(3dMDface-) based evaluation of the face in cleft lip infants. *Plastic surgery international,vol. 2012*.

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D Face Model for Pose and Illumination Invariant Face Recognition.*vol. 70*(80),pp. 90-96.

Popova, S., Lange, S., Probst, C., Gmel, G., & Rehm, J. (2017). Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis. *The Lancet Global Health,vol. 5*(3),pp. 290-299.

Rajamani, K. T., Styner, M. A., Talib, H., Zheng, G., Nolte, L. P., & Gonzalez Ballester, M. A. (2007). Statistical deformable bone models for robust 3D surface extrapolation from sparse data. *Medical Image Analysis,vol. 11*(2),pp. 99-109.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning* (pp. 63-71).

Romdhani, S., Blanz, V., & Vetter, T. (2002). *Face identification by fitting a 3d morphable model using linear shape and texture error functions.* Paper Presented at the European Conference on Computer Vision,pp. 3-19.

Romdhani, S., & Vetter, T. (2005). Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In (Vol. 2, pp. 986-993).

Roozen, S., Peters, G. J., Kok, G., Townend, D., Nijhuis, J., & Curfs, L. (2016). Worldwide Prevalence of Fetal Alcohol Spectrum Disorders: A Systematic Literature Review Including Meta-Analysis. *Alcoholism, Clinical and Experimental Research,vol. 40*(1),pp. 18-32.

Rusinkiewicz, S., & Levoy, M. (2001). *Efficient variants of the ICP algorithm.* Paper Presented at the Proceedings 3rd International Conference on 3-D Digital Imaging and Modeling,pp. 145-152.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., Little, R. E., Clarren, S. K., Dehaene, P., . . . Graham, J. M., Jr. (1997). Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology,vol. 56*(5),pp. 317-326.

Schonborn, S., Egger, B., Morel-Forster, A., & Vetter, T. (2017). Markov Chain Monte Carlo for Automated Face Image Analysis. *International Journal of Computer Vision,vol. 123*(2),pp. 160-183.

Schönborn, S., Egger, B., Morel-Forster, A., & Vetter, T. (2016). Markov Chain Monte Carlo for Automated Face Image Analysis. *International Journal of Computer Vision,vol. 123*(2),pp. 160-183.

Streissguth, A., Barr, H., Kogan, J., & Bookstein, F. (1996). Understanding the occurrence of secondary disabilities in clients with fetal alcohol syndrome (FAS) and fetal alcohol effects (FAE). *Final report to the Centers for Disease Control and Prevention (CDC)*,pp. 96-06.

Streissguth, A. P., Aase, J., Clarren, S., Randels, S., Ladue, R., & Smith, D. (1992). Fetal alcohol syndrome in adolescents and adults. *International Journal of Gynecology & Obstetrics,vol. 37*(4),pp. 332-332.

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Sampson, P. D., O'MALLEY, K., & Young, J. K. (2004). Risk factors for adverse life outcomes in fetal alcohol syndrome and fetal alcohol effects. *Journal of Developmental and Behavioral Pediatrics,vol. 25*(4),pp. 228-238.

Styner, M. A., Rajamani, K. T., Nolte, L.-P., Zsemlye, G., Székely, G., Taylor, C. J., & Davies, R. H. (2003). *Evaluation of 3D correspondence methods for model building.* Paper Presented at the Biennial International Conference on Information Processing in Medical Imaging,pp. 63-75.

Suttie, M., Foroud, T., Wetherill, L., Jacobson, J. L., Molteno, C. D., Meintjes, E. M., . . . Hammond, P. (2013). Facial dysmorphism across the fetal alcohol spectrum. *Pediatrics,vol. 131*(3),pp. 779-788.

Sykes, J. M., Suárez, G. A., Trevidic, P., Cotofana, S., & Moon, H. J. (2018). Applied Facial Anatomy. In B. Azizzadeh, M. R. Murphy, C. M. Johnson, G. G. Massry, & R. Fitzgerald (Eds.), *Master Techniques in Facial Rejuvenation* (pp. 6-14).

Szeliski, R., & Lavallee, S. (1994). Matching 3-D anatomical surfaces with non-rigid volumetric deformations. *Applications of computer vision in medical image processing*,pp. 22-25.

Takacs, B. (1998). Comparing face images using the modified Hausdorff distance. *Pattern Recognition,vol. 31*(12),pp. 1873-1881.

Tzou, C.-H., & Frey, M. (2011). *Evolution of 3D Surface Imaging Systems in Facial Plastic Surgery* (Vol. 19).

van Kaick, O., Zhang, H., Hamarneh, G., & Cohen-Or, D. (2011). A Survey on Shape Correspondence. *Computer Graphics Forum,vol. 30*(6),pp. 1681-1707.

Victor, J., Van Doninck, D., Labey, L., Innocenti, B., Parizel, P. M., & Bellemans, J. (2009). How precise can bony landmarks be determined on a CT scan of the knee? *Knee,vol. 16*(5),pp. 358-365.

Viergever, M. A., Maintz, J. A., Klein, S., Murphy, K., Staring, M., & Pluim, J. P. (2016). A survey of medical image registration–under review. *Medical Image Analysis,vol. 33*,pp. 140-144.

Wang, L., Zhang, Y., & Feng, J. (2005). On the Euclidean distance of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 27*(8),pp. 1334-1339.

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). *A 3D facial expression database for facial behavior research.* Paper Presented at the 7th International Conference on Automatic Face and Gesture Recognition (FGR06),pp. 211-216.

Zollhofer, M., Martinek, M., Greiner, G., Stamminger, M., & Sussmuth, J. (2011). Automatic reconstruction of personalized avatars from 3D face scans. *Computer Animation and Virtual Worlds,vol. 22*(2-3),pp. 195-202.